

GENE-BASED SYSTEMS APPROACH TO SIMULATE SOYBEAN GROWTH AND  
DEVELOPMENT AND APPLICATION TO IDEOTYPE DESIGN IN TARGET  
ENVIRONMENTS

By

CARLOS DANIEL MESSINA

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL  
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT  
OF THE REQUIREMENTS FOR THE DEGREE OF  
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2003

Copyright 2003

by

Carlos Daniel Messina

...to my grandmother Angela C. Untermann and to the memory of my grandparents

## ACKNOWLEDGMENTS

I wish to extend heartfelt thanks to my committee members and to the professors, graduate students and technicians who have contributed to this dissertation. It was a privilege and honor to work with them at, and for, the University of Florida.

Special thanks go to my advisor and mentor, Dr. James W. Jones. He gave me absolute freedom and independence of thought together with guidance and intellectual and economic support. His wisdom and expertise helped me focus and kept my research on the right track. He always had time to discuss research problems, to comment on manuscripts and to give me advice about my professional career. I could not have asked for a better advisor.

I would like to thank Dr. W. Graham, Dr. B. Hauser and Dr. J.T. Ritchie for their valuable advice as members of my committee. I wish to thank Dr. Boote for his valuable guidance during the conception of the research project, for reviewing draft documents, and for the pleasant and stimulating discussions we had about crop physiology and modeling. He also gave me the opportunity to teach plant physiology, for which I am very thankful.

This dissertation would not have been complete without the help of Dr. C.E. Vallejos. He had a critical role by teaching me genetics, molecular biology and laboratory techniques that allowed me to significantly improve this research. I am specially thankful for his thought-provoking questions and discussions, for his advice on grantsmanship, and for letting me participate in his research on beans.

I would like to acknowledge contribution of many people who provided valuable technical support, especially to Valerie Jones, Jean Thomas, Wayne Williams and Cheryl Porter. Dr. L. Bernard from the University of Illinois, provided the near-isogenic soybean lines I used to develop my model. Dr. R. Nelson, curator of the USDA germoplasm collection, provided soybean cultivars that I used to evaluate the model in performance variety trials. Dr. G. Podesta from the University of Miami provided me with high quality weather series for Argentine locations. Julio Dardanelli, Graciela. Magrin, Santiago. Meira and Maria Travasso kindly made available soil parameters and initial conditions, which I used to run the simulations in Argentina. Dr. Theodoros Mavromatis made available a variety trial database along with soil parameters and genetic coefficients for a set of varieties I used to test the model.

I wish to thank Dr. Roger Boerma and Dr. James Specht for their valuable comments while I formulated my dissertation project.

My experience at the University of Florida would have not been complete without the moments shared with Andrés Ferreyra and Fred Royce, and the interactions with Dr. Curtis Hannah, Dr. Charles Guy and Dr. Jane Luzar. Among the many things I have to acknowledge, I would like to thank them for helping me grow. Dr. Luzar, with her never-ending enthusiasm, helped Andres and me start the Town Hall and a mentoring and undergraduate research program at the Graduate Student Council.

To my family, I would like to express my gratitude for their dedication, support and love. I would like to express my dearest appreciation to my wife. Her endless love, friendship, encouragement and patience contributed greatly to this achievement.

## TABLE OF CONTENTS

	<u>Page</u>
ACKNOWLEDGMENTS .....	iv
LIST OF TABLES.....	ix
LIST OF FIGURES .....	xi
ABSTRACT.....	xiii
 CHAPTER	
1 General introduction .....	1
Statement of the Problem.....	1
The Bottom Up Approach .....	2
The Top Down Approach.....	5
Overall Objective and Organization of this Dissertation.....	8
 2 Genetic control of reproductive development and responses to photoperiod in soybean [ <i>Glycine max</i> L.].....	 11
Introduction.....	11
Materials and Methods .....	15
Field Experiments.....	18
Statistical Analyses.....	20
Results.....	22
Pod Addition Duration .....	23
Duration of Time Between Flowering and First Pod (R1-OPA).....	27
Duration between OPA and R5 .....	29
Final Pod Number and Pod Addition Duration .....	32
Duration between R5 and R7 .....	34
Discussion and Conclusions .....	35
 3 A gene-based approach to simulate soybean development and yield responses to the environment .....	 42
Introduction.....	42
Materials and Methods .....	45
Simulation of Soybean Development and Pod Addition.....	46
Data.....	49

Parameterization of CROPGRO-Soybean to Include Genetic Information .....	50
Parameter Estimation, Model Verification and Evaluation .....	52
Molecular Marker Length Polymorphisms Linked to <i>E</i> Loci and Cultivar Genotyping.....	54
DNA Extraction.....	55
Polymerase Chain Reaction (PCR) and PCR Product Separation.....	55
Predicting Time to Maturity of Public Cultivars of Soybean Grown in Variety Trials .....	56
Results and Discussion .....	57
Calibration and Evaluation of CROPGRO-Soybean.....	58
Soybean development during 2001 season .....	58
Simulation of pod addition duration and pod number in 2001 .....	60
Simulation of plant development effects on the dynamics of biomass accumulation, yield and harvest index.....	64
Estimating Genetic Coefficients from Genotypes .....	64
Evaluation of CROPGRO-Soybean with Genetic Coefficients Estimated from <i>E</i> Loci. Data collected during 2001 season. ....	66
Model Evaluation with Independent Data from 2002 .....	67
SSR Length Polymorphisms Linked to <i>E</i> Loci and Cultivar Genotyping.....	68
Predicting Soybean Yield and Maturity in Variety Trials.....	71
Conclusions.....	74
 4 Linking optimization algorithms and gene-based models for crop engineering in target environments .....	77
Introduction.....	77
Materials and Methods .....	80
Linking Crop Models and Optimization Algorithms to Assist Plant Breeding.....	80
Optimization with Adaptive Simulated Annealing .....	81
A Plant Breeding Metaphor for SA .....	86
Simulation of Crop Growth and Development.....	87
Linkage between SA and CROPGRO-Soybean.....	88
Identification of Traits Contributing to Yield Maximization in Target Environments .....	89
Genetic Base Breadth and Selection Pressure Effects on Yield Gains .....	91
Risks of Ignoring Epistatic and Pleiotropic Effects for Ideotype Design .....	92
Results.....	93
Convergence Towards a Global Maximum.....	93
Physiological Analysis of Parameter Dynamics.....	94
Identification of Traits Contributing to Yield Maximization in Target Environments .....	97
Ideotypes across a latitudinal range .....	97
Ideotypes across a water stress gradient.....	100
Genetic Base Breadth and Selection Pressure Effects on Yield Gains .....	102
Risks of Ignoring Epistatic and Pleiotropic Effects for Ideotype Design .....	105
Discussion and Conclusions .....	107
Linking SA and CROPGRO-Soybean for Ideotype Design.....	108

Optimal Plant Traits for Target Environments .....	110
Genetic Base Breadth and Crop Yields .....	114
Gene-based Models for Ideotype Design .....	115
5 SUMMARY AND CONCLUSIONS .....	118
LIST OF REFERENCES .....	123
BIOGRAPHICAL SKETCH .....	141

## LIST OF TABLES

<u>Table</u>	<u>page</u>
2-1. Clark and Harosoy near-isogenic lines used in this study .....	19
2-2. Analysis of variance for factors affecting soybean reproductive development, node number, and rate and duration of pod addition .....	25
2-3. <i>E</i> loci, growth habit and genetic background effects on pod addition duration (days) on different planting dates .....	26
2-4. <i>E</i> loci and <i>dt</i> control of the duration (days) between the onset of pod addition and R5 (beginning seed growth) on different planting dates .....	30
2-5. Effects of <i>E</i> loci and <i>Dt</i> on seed fill duration for different years, planting dates and genetic background .....	35
3-1. List of soybean near-isogenic lines used for model development and for molecular markers evaluation .....	49
3-2. Selected genetic coefficients in CROPGRO-Soybean controlling plant development in soybean, potential associations with <i>E</i> and <i>dt</i> loci, and variables used for parameter estimation .....	51
3-3. Comparative evaluation between CROPGRO-Soybean and CROPGRO-Soybean parameterized using <i>E</i> loci information using data collected during 2001. ....	58
3-4. Associations between <i>E</i> loci and genetic coefficients in CROPGRO-Soybean .....	65
3-5. Confirmation of CROPGRO-Soybean and CROPGRO- <i>E</i> loci to predict independent observations (2002) of soybean development. ....	68
3-6. Prediction errors in time to maturity and yield for soybean public varieties grown in eight locations in Illinois (1995-99) .....	72
4-1. Genetic coefficients, ranges of variation, and examples for four maturity groups ....	82
4-2. Associations between marker loci, soybean traits and genetic coefficients. ....	83
4-3. Soil parameters and soybean management for four locations in the Pampas .....	87

4-4. Simulated soybean growth and development for a set of genotypes across a latitudinal gradient .....	98
4- 5 Initial and optimized genetic coefficients, yield and yield components for Pilar, Pergamino and Balcarce. ....	99
4-6. Simulated soybean phenology for a set of genotypes under contrasting water stress environments in Santa Rosa. "Dry" 1945-55. "Wet" 1985-95.....	102
4-7. Initial and optimized genetic coefficients, yield and yield components for Santa Rosa under two contrasting environments: 1945-55 and 1985-95 .....	103
4-8. Maximum-yield sensitivity to variations in parameters $\Phi$ and $r_{\phi}$ .....	104

## LIST OF FIGURES

<u>Figure</u>	<u>page</u>
2-1. Soybean development during reproductive period, photoperiodic effects mediated by <i>E</i> and <i>dt</i> loci, and determination of yield components .....	12
2-2. Weather conditions during 2001 and 2002 at the research site .....	16
2-3. Soybean development during the reproductive period and relationship to pod number and total seed weight .....	17
2-4. Regression tree describing the data structure for duration of pod addition.....	24
2-5. Duration between flowering and the onset of pod addition as affected by planting date (early: May planting, late: July-August planting) and <i>E</i> loci.....	28
2-6. Relationship between pod addition duration and the onset of seed growth (as measured by time between OPA to R5) .....	31
2-7. Final pod number as a function of pod addition duration, genetic background, planting date, growth habit, and locus <i>E1</i> .....	33
3-1. Representation of functions used in CROPGRO-Soybean to account for shown photoperiod and temperature effects on soybean developmental phases and their relation to <i>E</i> loci (A), genetic coefficients for different phase duration (B), and yield components (C).....	48
3-2. Genetic variability on soybean development within the set of near-isogenic lines grown in Gainesville during 2001 and 2002.....	57
3-3. Relation between observed and predicted (A-F) time to flowering, (B-G) time to onset of pod addition, (C-H) time to last mainstem node (R5), (D-I) time to physiological maturity, and (E-J) pod addition duration .....	59
3-4. Observed and simulated pod number per plant for experiments conducted in Gainesville in two planting dates during 2001 .....	61
3-5. Relationship between R5 and the duration of pod addition.....	62
3-6. Simulated biomass accumulation for the near-isogenic lines L98- 2064 (□) and L92- 21 (○) grown in Gainesville in 2001.....	63

3-7. SSR length polymorphisms linked to <i>E</i> loci.....	69
3-8. <i>E</i> loci genotypes for a set of soybean public cultivar.....	70
3-9. Simulated and observed time to maturity (day of the year) and yield (kg ha <sup>-1</sup> ) of a seven soybean public varieties grown at eight locations and five years in Illinois (1995-99) .....	72
4-1. Representation of the linkage between SA algorithm and CROPGRO-Soybean. ....	88
4-2. Case study location in the Argentine Pampas (A) and water stress index dynamics during the growing season (B).....	91
4-3. Evolution of simulated soybean yields in Balcarce during SA optimization .....	94
4-4. Genetic coefficient dynamics during SA optimization for Balcarce .....	95
4-5. Simulated biomass, water stress index and leaf area index (LAI) dynamics for probe (P), initial (I) and optimized (O) soybean genotypes for Santa Rosa in contrasting environments of water availability .....	101
4-6. Simulated total and seed biomass for SA optimized genotypes grown in Balcarce (A) and Pergamino (B).....	105
4-7. Probabilities of exceedence of yield simulated for probe, optimized genotypes and <i>E</i> - loci optimized genotypes in Balcarce (BALC) and Pergamino (PERG).....	107

Abstract of Dissertation Presented to the Graduate School  
of the University of Florida in Partial Fulfillment of the  
Requirements for the Degree of Doctor of Philosophy

GENE-BASED SYSTEMS APPROACH TO SIMULATE SOYBEAN GROWTH AND  
DEVELOPMENT AND APPLICATION TO IDEOTYPE DESIGN IN TARGET  
ENVIRONMENTS

By

Carlos Daniel Messina

December 2003

Chair: James W. Jones

Major Department: Agricultural and Biological Engineering

Crop yields must increase to satisfy an increasing food demand. Plant breeding and improved crop management will constitute the backbone for breaking productivity constraints. Rapid advances in molecular biology promise to radically change plant genetic improvement. However, we need methods to bridge the gap between genes and crop performance, to predict crop responses to environmental conditions and management, and to design predictable phenotypes.

Crop models, software programs that imitate plant growth and development, have the potential to become powerful genetic engineering tools. Paradoxically, model parameters that characterize genotypic differences are phenotypic in nature. If these can become functions of loci, we can establish a bridge between genetics, crop biology, and crop and environmental management. This dissertation develops, tests, and demonstrates an approach to tailor a crop model to the genetic makeup of the crop for ideotype design for target environments.

Using soybean as a model organism, a set of *E* loci that control reproductive development was studied using 48 near-isogenic lines. New functions were assigned to the *E5* locus and other *E* loci that control reproductive duration and pod number determination. These experimental results were used to develop linear models to predict crop model parameters from *E* loci alleles bridging the gap between genetics and integrated crop physiology. For the first time, this kind of approach was tested for its ability to predict growth and development in commercial varieties. The gene-based model predicted 75% of the variance in the time to maturity and 54% of the yield variance in variety trials conducted in Illinois. Gene-based approaches can thus reduce or replace expensive and time-consuming experimentation for model parameterization. A phenotype reverse-engineering method was implemented by coupling the gene-based model to a simulated annealing optimization algorithm. The new method was used to design ideotypes for target environments in Argentina. The coupled model found ideotypes yielding at least 40% more than actual varieties grown in the region. Although more research is needed to fully parameterize the soybean model, it was shown that there is great potential for decreasing model parameterization requirements, and for designing ideotype for food production systems.

## CHAPTER 1 GENERAL INTRODUCTION

### **Statement of the Problem**

Awareness about world hunger is leading society to pay more attention to the need for significant crop improvement in poor countries. If crop yields do not increase further in productive regions, expansion to marginal lands will be necessary. This will require crop varieties that withstand drought, impoverished soils, heat and cold stresses, and disease (Kennedy, 2003).

Plant breeding has produced germplasm with increased yield; the potential yields of most crops have increased continuously since the 1930s (Lee, 1995). However, Sinclair (1993) suggests that there are only marginal opportunities for further genetic improvements in crop yield potential. Development of high yielding cultivars and varieties adapted to increasingly diverse environments thus presents a great challenge to conventional plant breeding (Leach et al., 2002). Alternative strategies may be required to design plant ideotypes suitable for specific target environments.

Recent advances in the sequencing of plant genomes (The Arabidopsis Genome Initiative, 2000; Yu et al., 2002; Goff et al., 2002), plant functional genomics, and genetic engineering technologies promise to radically change plant genetic improvement (Somerville and Somerville, 1999; Cooper et al., 2002; Chapman et al., 2003; Lee, 1995; Ronald and Leung, 2002). However, the realization of this goal will depend on the development of methods and tools that will allow us to gain an intrinsic understanding of

complex biological systems. Only then may we be able to make rational changes for the production of best-adapted phenotypes.

Systems approaches are so far the best paradigm to study, understand and manipulate complex systems (Jones and Luyten, 1998; Sandquist, 1985; Kitano, 2002; Csete and Doyle, 2002). Although systems approaches were widely used to understand plant systems (Jones et al., 2003; Keating et al., 2003; van Ittersum et al., 2003), only recently have there been efforts to use molecular level knowledge to simulate organs (Noble, 2002), embryos (Davidson et al., 2002), and specific plant traits (White and Hoogenboom, 1991; Raymond et al., 2003; Yin et al., 2003; Stewart et al., 2003). We now have the opportunity to design plant ideotypes starting from the very basic biological principles at the molecular level. The development and application of mathematical concepts and systems approaches to uncover principles underlying biology at the molecular, cellular and organism levels are emerging under the name of computational systems biology (Kitano, 2002; Ideker et al., 2001).

### **The Bottom Up Approach**

The current ability to perturb an organism (Valenzuela et al., 2003) and monitor whole-genome gene expression (Brown and Botstein, 1999; Duggan et al., 1999; Lockhart and Winzeler, 2000), global protein accumulation (Ghaemmghami et al., 2003; Braun and LaBaer, 2003), protein modification dynamics in the cell (Raghothama and Pandey, 2003), and large numbers of metabolite concentrations (Weckwerth, 2003) opens an unprecedented opportunity to understand, simulate and study the dynamics of organisms as systems.

However, we may overestimate the potential of these technologies and knowledge to assist genetic improvement in crops. One may envision that by measuring whole-

genome gene expression under selected conditions in a cultivar of interest we will be able to identify genes that regulate the trait of interest and engineer plants accordingly (Ronald and Leung, 2002). Alternatively, we can think of complex models that will use information about each gene product to simulate the growth and development of a plant under any environmental conditions (Somerville and Dangl, 2000). A model of this kind, with at least 25000 genes or 100000 proteins as state variables (The Arabidopsis Genome Initiative, 2000), would probably be as complex as the real organism. This contradicts the underlying idea of a model serving as a simplified representation of reality, but its contribution as a descriptive tool is undeniable. Whether a 25000 state-variable model could improve our understanding of an organism beyond the lessons that can be learned by manipulating the organism itself remains unanswered. The capability of such a model to predict the organism's overall behavior is questionable.

Bottom-up approaches to modeling living organisms face several challenges. The first arises from the complexity of biological systems and their signaling mechanisms (Weng et al., 1999). Genome-wide modeling efforts from yeasts to humans include an array of methods including principal components (Holter et al., 2001; Huang et al., 2003; Alter et al., 2000), self-organizing feature maps (Maleck et al., 2000), clustering (Schenk et al., 2000) and network analysis (Brazhnik et al., 2002; Stark et al., 2003; Perrin et al., 2003; Friedman, 2003; Tamada et al., 2003). Most of these studies sought to identify gene expression patterns and represent them as gene networks. Others claimed their model had predictive abilities (Holter et al., 2001; Huang et al., 2003; Stark et al., 2003), but the power of these statistically-based approaches under different environmental conditions remains to be proven.

Network analysis and the identification of coordinated patterns of gene expression led to the realization that most gene products act in complexes, the molecules of which are densely connected with each other, but sparsely connected with the rest of the network (Spirin and Mirny, 2003; Rives and Galitski, 2003). These findings support an early proposition (Hartwell et al., 1999) that suggested a shift from molecular to modular cell biology. The “modular framework” (Jones, 1998; Acock and Reynolds, 1997; Reynolds and Acock, 1997) implicit in this network-of-complexes metaphor allows us to model cell systems at a mechanistic level using a reduced number of variables and assays (Kholodenko et al., 2002). Mathematical approaches for the simulation of the dynamics of these subsystems can be found elsewhere (Smolen et al., 2000; Ideker et al., 2001; Gilman and Arkin, 2002; McAdams and Arkin, 1998)

Many models targeted to simulate small sub-networks within cells, such as circadian rhythms (Gonze et al., 2002; Leloup and Goldbeter, 2000), partial signal transductions (Sachs et al., 2002; Schoeberl et al., 2002), light- and carbon-signaling pathways in plants (Thum et al., 2003), and cell cycles (McAdams and Arkin, 1998), successfully reproduced observations. Furthermore, emergent properties such as integration of signals across multiple time scales and self-sustaining feedback loops of some networks of biological signaling pathways were demonstrated (Bhalla and Iyengar, 1999; Leloup and Goldbeter, 2000).

Integrating all these subsystems to simulate organs or whole organisms is a colossal challenge, even when using modular approaches at the cell level. Many subsystems differ by orders of magnitude in terms of time and spatial scales. Many processes are sequential, beginning with gene transcription and translation, followed by protein

synthesis, intracellular transport of proteins, biochemical synthesis of metabolites, and ending with long distant (intercellular) transport of molecules. Investigations on protein networks and dynamics, signal transduction, regulatory mechanisms regulating phenotypic plasticity and cell and tissue communication are in their infancy, however. The lack of knowledge as today and the complexity intrinsic to a model of multiple parameters ( $\sim 10^5$  proteins), each one having attendant uncertainty, can propagate errors, reducing the ability of a model to predict phenotypes (Thornley and Johnson, 1990).

Advances in molecular biology will allow us to develop the technologies that change plant genetic improvement as we now it today. However, bottom up approaches face several challenges that made this path unviable today for predicting phenotypes from genotypes limiting its application in plant breeding.

### **The Top Down Approach**

Dynamic simulation biophysical models of plant growth and development are based on the state-variable approach (Jones et al., 2003; Keating et al., 2003; van Ittersum et al., 2003). State variables describe the conditions of each component of the system; many represent tangible quantities such as leaf mass (Jones and Luyten, 1998). Together with environmental variables, these state variables determine how the plant responds to ontogeny and environmental conditions. A subset of state variables is used to storing information rather than mass; these are mathematical artifacts of the simulation model that mimic analog information systems acting in the plant (Jones and Luyten, 1998), such as hormone-mediated messaging (Buchanan et al., 2000), and more recently microRNA-based communication (Aukerman and Sakai, 2003).

The absence of hormones or detailed communication mechanisms of any form in existing crop models is not fortuitous. Modelers have used three main arguments to

explain modeling approaches without hormone action (de Wit and Penning de Vries, 1983). The first argument builds upon the lack of knowledge about how hormones are produced and de-activated or degraded, about their rates of translocation, the kinetics of their action, and the population variation in sensitivity of plant cells to hormones (Bradford and Trewavas, 1994). An extension of this argument is the lack of experiments adequate for modeling. As discussed above, the strength of this argument is weakening due to the rapid progress in molecular biology.

The second argument focuses on the role of the hormonal system as a communication system. If we are able to mimic the communication system by using proxy variables, then there is no need to include detailed components in the system that may contribute to model instabilities rather than to gain understanding about the functionality of the crop. The very strong underlying assumption in this approach is that all of the plant's biochemical processes are functioning "properly" in the experiments used to develop the model. Studies on natural variation in light sensitivity of *Arabidopsis* demonstrate that the meaning of the word "properly" can change with environmental conditions, however. For example, a single amino acid substitution in cryptochrome 2 and phytochrome A in natural populations of *A. thaliana* reduced the functionality of these proteins but conferred tremendous competitive advantages when grown in low latitudes (Maloof et al., 2001; El-Assal et al., 2001) at high irradiance.

The high speed of hormonal-balance processes relative to the rate of change in most state variables in a crop model set the basis for arguing that hormones are not essential in crop models (de Wit and Penning de Vries, 1983). If hormones reach equilibrium within a time span lower than the time step used for numerical integration in

the model, then it is possible to relate the cause driving the hormonal processes directly to the physiological process affected by the hormones. An extension of the argument is that there is no information in the model to drive such fast processes and even less data for their evaluation. A generalization of de Wit and Penning de Vries (1983) ideas regarding hormone-related processes is that not more than two or three hierarchical levels should be simulated in crop models.

Although taking important signaling mechanisms for granted (e.g., those mediated by phytochromes) led to failure in simulating adequately the leaf area dynamics in wheat (Meinke et al. 1998), the analyses presented by de Wit and Penning de Vries (1983) demonstrate that in some cases it is not worth attempting to integrate models of different hierarchical levels if system predictability is the primary aim.

Accurate model prediction of phenotypes for a given genotype in different environments is required for crop models to be useful tools in agronomy and plant breeding. Genetic differences in existing crop models (e.g., Jones et al., 2003; Keating et al., 2003; van Ittersum et al., 2003) are represented by cultivar-specific parameters, which are paradoxically phenotypic in nature. Using this representation of genetic differences among cultivars, crop models proven accurate predicting genetic differences and interactions with the environment (Mavromatis et al., 2001; Boote et al., 2001). Because of the phenotypic nature of these parameters, it is uncertain, however, how well epistatic and pleiotropic effects are being represented, what are the processes at the molecular level that these parameters represent, and what are the causes underlying the variations in these cultivar-specific parameters. These uncertainties limit the applicability of existing crop models in plant breeding and plant biology.

Existing crop models, however, can provide us with the biophysical and physiological framework as a well-organized system. We can use this framework to develop top-down approaches, which can bridge the gap between genotypes and plant phenotypes, by making cultivar specific parameters functions of qualitative and quantitative loci. The top-down approach could be further expanded by replacing model components and parameters by modules that simulate gene networks, interactions between gene products and other metabolites (White and Hoogenboom, 2003).

### **Overall Objective and Organization of this Dissertation**

This research contributes to the development of the emerging discipline of systems biology, with emphasis on the simulation of plant growth and development for ideotype and food production systems design. The overall objective of this dissertation is to develop and test a systems approach for ideotype design based on previously characterized alleles at selected loci in soybean.

The research is organized into three interconnected core chapters. Each chapter is self-contained and addresses specific objectives. Specific background information is provided in the introduction of each chapter. The Materials and Methods and Discussion sections in Chapters 2 and 3 refer to previous chapters. Each chapter builds upon the previous one. It is recommended that they be read in sequence to fully understand the methodology. The organization and specific objectives of individual chapters are presented below.

Chapter 2 uses soybean as a model organism to study the genetic control of response to photoperiod mediated by *dt* and *E* loci during the reproductive period, and to evaluate their effects on fruit number. Previous research reported the effects of *E* loci on time to flowering and maturity (Cober et al., 2001; Cober et al., 1996; McBlain et al.,

1987). However, we had incomplete knowledge about the effects of *E* loci on critical sub-periods of the reproductive development of soybean. A field experiment was conducted to test the following hypotheses that:

- The *dt* and *E* loci regulate the duration of the following periods: a) from first flower to first pod; b) pod addition; c) seed filling; and d) from first flower to the onset of seed development.
- *E* loci regulate pod number by affecting the rate of pod addition.
- *E* loci regulate duration of pod addition by regulating the onset of seed development.

New experimental evidence of genetic control of pod addition duration in response to photoperiod in soybean and *E* loci effects on pod number is provided. The experimental results and data collected in this chapter are critical for the development of the model described in Chapter 3. These experiments were necessary because experimental manipulations of the environment in previous experiments described in the literature were inadequate for the development of a model aimed at predicting soybean growth and development at a field scale.

Chapter 3 develops and evaluates a gene-based biophysical model that simulates soybean growth and development using experimental data generated in Chapter 2. The model is further evaluated for its ability to predict the soybean development and yield results of a variety trial conducted in Illinois. This evaluation is the first one conducted on a model of this kind. Informative microsatellites closely linked to *E* loci were identified and used to determine the allelic combinations for each cultivar.

Chapter 4 describes a methodology for ideotype design for target environments that tailors crop simulation models and a global optimization algorithm (simulated annealing), for which a new metaphor is introduced. The method is evaluated by its capability to

identify traits contributing to yield maximization in target environments, and is demonstrated in two applications. One application studies the effects of breadth of genetic base and selection pressure on yield gains, and the other studies the risks of ignoring epistatic and pleiotropic effects in ideotype design.

Chapter 5 presents a summary and an integration of the conclusions.

## CHAPTER 2

### GENETIC CONTROL OF REPRODUCTIVE DEVELOPMENT AND RESPONSES TO PHOTOPERIOD IN SOYBEAN [*Glycine max* L.]

#### Introduction

Soybean yield is the result of many complex interactions among the genetic makeup of the crop, physiological processes and the environment throughout crop development. The realization of yield potential and how fruit number and weight determine it depend upon the partitioning of the reproductive period into fruit addition and fruit growth phases. Fruit number is the main determinant of final yield (Shibles et al., 1975; Egli, 1998), while individual seed weight and individual seed growth rate (ISGR) generally show weak or no correlations to final yield (Egli, 1998, Guffy et al., 1991). In order to accurately predict yield and to design successful breeding strategies, we must first identify and characterize the physiological and genetic determinants of fruit number, and the environmental factors that affect them.

The current hypothesis to explain the determination of seed number is that the number of seeds is set such that the summation of ISGR across fruit cohorts reaches an equilibrium with the ability of the soybean canopy to supply assimilate to support fruit growth (Egli, 1998; Wardlaw, 1990) (Fig.2-1). Once the growing fruit is supplied with a minimum accumulation of assimilates, its growth will continue (Charles-Edwards, 1984a,b; Charles-Edwards and Beech, 1984). Several studies provide evidence that supports this hypothesis, showing that final fruit number correlates with the intercepted

radiation and crop growth rate during the reproductive period (Kantolic and Slafer, 2001; Egli and Bruening, 2000; Egli et al. 1985; Jiang and Egli, 1993).

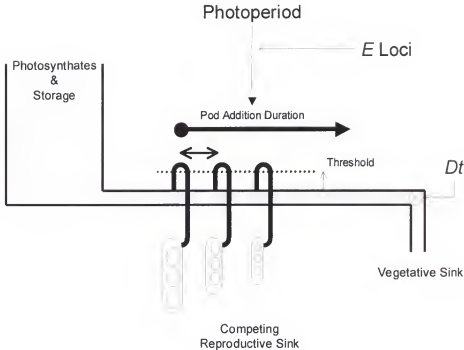


Figure 2-1. Soybean development during reproductive period, photoperiodic effects mediated by *E* and *dt* loci, and determination of yield components. Adapted from Wardlaw (1990)

Plants have evolved mechanisms to perceive and respond to environmental cues to maximize their adaptation to the environment. Soybean is a short day plant. Photoperiod controls soybean development and resource allocation, regulating the onset and duration of the period of addition and growth of reproductive structures. Long photoperiods delay the time to flowering (Thomas and Vince Prue, 1997) and maturity (Johnson et al., 1960). During the reproductive period, long photoperiods extend the duration of flowering (Summerfield et al, 1998) but decrease the rate of flower differentiation (Zhang et al., 2001; Thomas and Raper, 1984; Board and Settini, 1988; Fisher, 1963) and the reproductive efficiency, the ratio between the number of pods and flowers (Fisher, 1963;

van Schaick and Probst, 1958). Long days also delay the onset of pod addition (Fisher, 1963; Johnson et al., 1960) and pod growth (Wilcox et al., 1995; Board and Settimi, 1986), and extend pod addition (Kantolick and Slafer, 2001) and seed filling periods (Guffy et al. 1991; Raper and Thomas, 1978). These photoperiodic effects on crop development translate into changes in yield and yield components. The extension of pod addition duration correlates with an increased number of seeds in response to higher intercepted radiation (Kantolick and Slafer, 2001). Similarly, an extended time period before pod filling increases seed number (Egli and Bruening, 2000) and is associated with a larger number of nodes. Assuming co-regulation of the onset of pod set and seed set (Mavromatis et al. 2001), the Egli and Bruening results could be interpreted as the consequence of an extension in the duration of the pod addition period and a delay in the sink development. The extension of seed filling period with longer photoperiods can be the consequence of a longer pod addition duration, an increased duration of the individual seed filling period, or both (Fig.2-1).

The genetic controls of some of the aforementioned soybean responses to photoperiod were characterized. Growth habit is controlled at the *dt* locus, where the dominant allele *Dt* conditions indeterminate growth and the recessive allele *dt* causes determinate growth. This loci (*Dt*) delays the change of the apex from a vegetative to reproductive state in response to longer photoperiod, thereby regulating the generation of leaf area, leaf area expansion, resource allocation and branching patterns (Wilcox et al., 1995). A set of independent *E* loci regulate time to flowering and maturity responses to photoperiod (Cober et al., 1996). The dominant alleles at *E2*, *E3*, *E4* and *E5* lengthen the duration of the reproductive phase under long photoperiods (Bernard, 1971; McBlain and

Bernard, 1987; Cober et al., 1996), while the dominant allele at the *E1* locus hastens reproductive development (McBlain et al. 1987). Recent data suggest that *E1* delays the onset and duration of seed fill and maturity under extended daylength (Curtis et al., 2000) similar to the action of dominant alleles at the *E4* (Saindon et al., 1989; Curtis et al., 2000) and *E3* loci (Curtis et al., 2000). Although the effects of these alleles are similar to *Dt* effects on seed filling rate and yield, they may differ in their mechanisms of action. While the allele *Dt* only affects the number of nodes, hence potentially extending the duration of pod addition and pod number, dominant alleles at the *E* loci can also increase crop radiation use efficiency (Ellis et al., 2000), possibly the main mechanism by which they affect seed filling rate. However, reported differences in radiation use efficiency between near isogenic lines (NILs) may require alternate interpretations if they arise due to differences in plant composition, specific leaf area and harvest index not taken into account by Ellis et al. (2000). An alternative hypothesis that can explain the effects of *E* loci on seed filling rate is that these loci affect the duration of pod addition, hence affecting seed number and seed filling rate.

Recent work demonstrates that *E1*, *E2* and *E3* regulate the length of the flowering period as a function of photoperiod (Summerfield et al., 1998). Positive epistasis has been detected between *E1*, *E2* and *E3* (Asumadu et al., 1998). Effects of *E2* and *E3* on flowering duration are strongly enhanced by the presence of *E1*. These results are in contrast with previous findings by McBlain et al. (1987). The *E* loci-mediated extension of flowering period suggests that they can regulate the duration of pod addition. This condition is not sufficient, however, since flower and small embryo abortion is high early and late in the reproductive period (Tischner et al., 2003). Flower and early pod shedding

is strongly associated with duration of flowering (van Schaick and Probst, 1958; Wilcox et al., 1995).

Because of the high correlation between fruit number and yield (Shibles et al., 1975; Egli, 1998), it is of great importance for physiologists, modelers and breeders to understand the genetic regulation of the reproductive period, in particular the genetic control of the duration of pod addition, which is a critical period for yield determination. The objectives of this dissertation are to study the genetic control of soybean response to photoperiod mediated by *dt* and *E* loci during the reproductive period, and to evaluate their effects on fruit number. We tested the hypothesis that the *dt* and *E* loci regulate the duration of the following periods: a) from first flower to first pod, b) pod addition, c) seed filling, and, d) from first flower to the onset of seed development. We also tested the hypothesis that *E* loci regulate pod number by affecting the rate of pod addition, and that *E* loci regulate duration of pod addition by regulating the onset of seed development.

### **Materials and Methods**

We studied the genetic regulation of soybean reproductive development, the duration of the critical period of pod addition, and its impact on the determination of pod number. The approach used a set of near isogenic lines carrying different allelic combinations at several *E* loci, which are known to delay maturity, and at the *dt* locus which regulates growth habit in response to photoperiod. The isolines were planted on two different dates to exploit the differences in photoperiod during the growth cycle (Fig. 2-2), and evaluate the effect of the loci under study. We tested the hypotheses in two genetic backgrounds to assess the general validity of the results and to test whether the *E* and *Dt* loci effects are dependent upon the presence of other genes.

This dissertation investigates the genetic regulation of subphases during the reproductive period (Fig. 2-3). The first subphase begins when the first flower becomes visible (R1) and ends at the onset of pod addition (OPA). We define here OPA as the time when 50% of the plants have a pod larger than 5 mm anywhere on the plant. During this phase, the reproductive period is initiated but no seeds are set. In the context of the conceptual model in Figure 2-1, only a minor fraction of assimilates are allocated to reproductive sinks during this phase. The duration of this phase can affect the duration of pod addition, seed fill duration and partitioning of assimilates to reproductive sinks affecting the final seed harvest index.

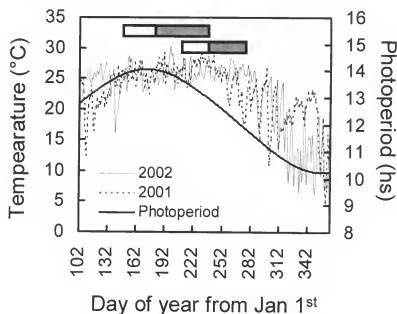


Figure 2-2. Weather conditions during 2001 and 2002 at the research site. Measurements were taken 40 m from the experimental plots. Growing seasons for two planting dates are shown indicating pre and post-flowering periods.

The duration of the pod addition phase defines a critical period for addition of reproductive sinks. Previous work estimated this window as the period beginning at R3 (see Fehr and Caviness (1977) for definitions of R stages in soybean) and ending at R6. This definition, however, has some limitations. First, OPA begins before R3, particularly

in indeterminate soybean. Between R3 and R6, seeds start growing (Fig.2-3), decreasing the amount of assimilates to set new sinks (Fig.2-1); therefore, R6 may not be related to the end of pod addition and may not be consistent across locations, years and planting dates provided that seed growth rate varies with these factors. To prevent this limitation we used OPA instead of R3 to improve the estimation of pod addition duration. We define the end of pod addition as the time when 50% of the plants have one abscising pod larger than 5 mm anywhere on the plant. Typically, these late added pods start but fail to be carried; thus this event serves as an indicator of the time when the maximum capacity of the canopy to support pods has been achieved. The difference between these two events provides an accurate characterization of the duration of the critical window for pod addition.

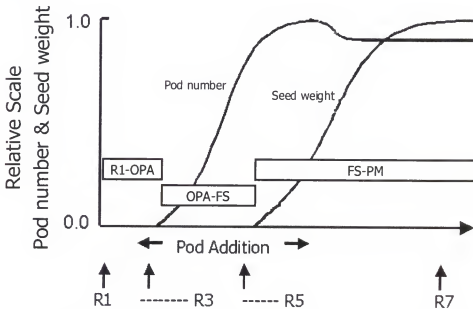


Figure 2-3. Soybean development during the reproductive period and relationship to pod number and total seed weight. R1: first flower visible anywhere in the main stem, OPA: onset of pod addition, FS: first seed visible anywhere in the plant, R5: first seed visible in any pod in the upper four nodes, R7: estimator of physiological maturity (PM)

The time between OPA and first seed (FS) defines a period of fruit addition with low or no competition between reproductive sinks for assimilates. Under the conceptual model used in this dissertation (Fig.2-1), the end of pod addition must occur after FS, and the duration between the onset of pod addition and FS should be associated with the duration of pod addition. If there is independent genetic regulation of the phases: R1-OPA and OPA-FS, then the duration of pod addition can be extended by reducing the lag time between R1 and OPA without causing an early onset of FS.

This paper uses R5 as an estimator of FS. R5 is an accurate estimator of FS in determinate growth habit NILs (*dt*), however, it is a less accurate estimator of FS for indeterminate (*Dt*) NILs since R5 is measured on the top four nodes and FS can be first set at any node in the plant.

There is consensus that seed fill duration can be defined as the duration between R5 and R7 since it is highly correlated with the effective seed filling period (Nelson, 1986 ; Guffy et al., 1991). Soybean breeders have shown that this period correlates with seed yield (Smith and Nelson, 1986; Boerma and Ashley, 1988) and that it is a heritable trait (Pfeiffer and Egli, 1988). Therefore, it is considered a good criterion for selection in plant breeding for higher yield (Nelson, 1986).

### **Field Experiments**

The set of NILs (Table 2-1) was grown in the field at the University of Florida, Gainesville, Florida, USA (29.630° N ; 82.370° W), in 2001 and 2002. Soybeans were planted on May 22<sup>nd</sup> and July 25<sup>th</sup> in 2001 and on May 23<sup>rd</sup> and August 7<sup>th</sup> in 2002. Plots were single rows, 3 m long in 2001 and 1.5 m long in 2002. Plots were hand planted with 20 seed per meter of row and row spacing was 56 cm.

Table 2-1. Clark and Harosoy near-isogenic lines used in this study

<i>Dt</i> and <i>E</i> loci	Genetic Background	
	Harosoy	Clark
<i>dt e1 e2 e3 e4 e5 e7</i>	OT94-43	
<i>dt e1 e2 e3 e4 e5 E7</i>	OT89-6	
<i>dt e1 e2 e3 E4 e5 E7</i>	OT94-37	L80-5882
<i>dt e1 e2 E3 e4 e5 E7</i>	OT94-39	
<i>dt e1 e2 E3 E4 e5 E7</i>	L67-153	L65-778
<i>dt e1 E2 e3 E4 e5 E7</i>		L63-3270
<i>dt e1 E2 E3 E4 e5 E7</i>		L63-3016
<i>dt E1 e2 e3 e4 e5 E7</i>	OT94-51	
<i>dt E1 e2 e3 e4 e5 E7</i>	OT94-49	
<i>dt E1 e2 e3 E4 e5 E7</i>	L74-102	L80-5879
<i>dt E1 e2 E3 E4 e5 E7</i>	L71-1116	L66-531
<i>dt E1 E2 e3 E4 e5 E7</i>		L76-865
<i>Dt E1 E2 E3 E4 e5 E7</i>		L66-546
<i>Dt e1 e2 e3 e4 e5 e7</i>	OT94-47	
<i>Dt e1 e2 e3 e4 e5 E7</i>	OT89-5	
<i>Dt e1 e2 e3 E4 e5 E7</i>	L62-667	L71-920
<i>Dt e1 e2 e3 E4 E5 E7</i>	L84-337	L97-2076
<i>Dt e1 e2 E3 e4 e5 E7</i>	OT94-41	L92-21
<i>Dt e1 e2 E3 E4 e5 E7</i>	Harosoy	L63-3117
<i>Dt e1 e2 E3 E4 E5 E7</i>	L64-4830	L94-1110
<i>Dte1 E2 e3 e4 e5 e7</i>	OT99-17	
<i>Dt e1 E2 e3 E4 e5 E7</i>	L84-307	L63-2404
<i>Dt e1 E2 E3 E4 e5 E7</i>	L64-4584	Clark
<i>Dt e1 E2 E3 E4 E5 E7</i>	L74-66	L92-1195
<i>Dt E1 e2 e3 e4 e5 E7</i>	OT93-28	
<i>DtTE1 e2 e3 e4 e5 E7</i>	OT93-26	
<i>Dt E1 e2 e3 E4 e5 E7</i>	L71-802	L80-5914
<i>Dt E1 e2 E3 E4 e5 E7</i>	L67-2324	L66-432
<i>Dt E1 e2 E3 E4 E5 E7</i>	L71L-3015	L97-4081
<i>Dt E1 E2 e3 E4 e5 E7</i>		L74-441
<i>Dt E1 E2 E3 E4 e5 E7</i>	L71L-3004	L65-3366
<i>DtE1 E2 E3 E4 E5 E7</i>		L98-2064

Weeds were controlled by the application of a pre-emerge herbicide application of 1.0 g m<sup>-2</sup> of Pendimethalin [*N*-(1-ethylpropyl)-3,4-dimethyl-2,6-dinitrobenzenamine]. During the growing season, weeds were also removed by hand. Pests and diseases were controlled by applications of 1.17 ml m<sup>-2</sup> of Daconil [tetrachloroisophthalonitrile] and 0.07 ml m<sup>-2</sup> of Permethrin [(3-phenoxyphenyl)methyl 3-(2,2-dichloroethenyl)-2,2-dimethylcyclopropanecarboxylate] applied as necessary, varying between seasons and planting dates. Fertilizer (14:14:14) was applied in bands at a rate of 16.8 g m<sup>-2</sup>, when the

plants reached the 4-leaf stage. Plants were grown under irrigation. Within planting dates and years, NILs were planted in two complete randomized blocks. Observations of R1, OPA, R5, end pod addition and R7 were taken between three and four times a week. Pod numbers and nodes on branches were measured after harvesting.

### Statistical Analyses

To test the hypothesis stated about the effects of *dt* and *E* loci on plant development and rate of pod addition, we used a mixed linear model in S-Plus (Pinheiro and Bates, 2000). Mixed linear models are frequently used to model grouped data, because they model flexibly the within-group correlation often present in this kind of data. A mixed linear model for the group *i* (defined in this study by year, planting date and block) has the general form,

$$Y_i = X_i\beta + Z_i b_i + \epsilon_i \quad i = 1, \dots, 8 \quad (1)$$

where  $Y_i$  is the response variable vector for group *i* (see below for grouping description),  $\beta$  is the vector of unknown fixed effects coefficients,  $b_i$  encode unknown random effects for group *i*,  $X_i$  and  $Z_i$  are known fixed effects and random effects regressor matrices, and  $\epsilon_i$  is the unknown within group error vector.

The data were analyzed as a nested design, where genotypes (NILs) were nested within planting date and year in a complete randomized block design. The experimental design determined eight groups (*i*) indicated in equation 1 (2 years x 2 planting dates x 2 blocks). Year, planting date, genetic background, *E* and *dt* loci and their interactions were fixed effects coded as columns in  $X_i$  (eq. 1). Each row in  $X_i$  corresponds to one NIL. Therefore,  $X_i$  has dimensions of 48 (NILs) by *n*, where *n* varies with the number of variables included in the model (see below for the procedure for selection of variables).

For example, if we are testing for main effects of loci  $E1$ ,  $E2$  and  $E3$  on R1-OPA, then  $\mathbf{X}_i$  is a  $48 \times 3$  matrix (see Table 2-2 for variables included in the model). Then, the vector  $\beta$  has dimension  $3 \times 1$  coding for the estimated coefficients for  $E1$ ,  $E2$  and  $E3$ . Matrix  $\mathbf{X}_i$  has entries of 1 and -1 coding for example for dominant (1) and recessive (-1) alleles.

Blocks nested within planting date and year were random effects encoded in  $\mathbf{Z}_i$  (eq.1) and form eight groups. The dimension of  $\mathbf{Z}_i$  is  $48 \times 1$  since block is the only random variable. In this case,  $b_i$  is a scalar encoding for block effects. Model parameters,  $\beta$ ,  $b_i$  and  $\alpha_i$  were estimated using maximum likelihood.

To guide the process of variable selection to include in the mixed-linear models, we first conducted a search for variables with high discrimination power using classification and regression trees (CART) (Venables and Ripley, 1997). Predictor variables included in the CART model were year, planting date, genetic background,  $E$  and  $dt$  loci. Response variables were R1-OPA, OPA-R5, R5-R7, pod addition duration, and pod number per unit day of pod addition.

The CART technique searches the variable space for those predictor variables that produce the best binary partition of the response variable. This process is repeated for successive partitions such that the split of a node produces a reduction in the deviance of the leaves (see below) relative to the node. The outcome of this procedure is a tree, a set of hierarchical partitions. To prevent over-fitting, we used cross validation to identify the number of partitions that minimize for the whole tree, the cost-complexity measure (Venables and Ripley, 1997),

$$D_\alpha = \sum D_j + \alpha n \quad j = 1, \dots, n$$

where  $D$  is the deviance of the response variable,

$$D_j = \sum (y_{ij} - y_m)^2 \quad i = 1, \dots, m$$

and  $\alpha$  is a parameter,  $n$  is the size of the tree,  $y_{ij}$  denotes observation  $i$  in leaf  $j$  and  $y_m$  is the mean of the observations for leaf  $j$ . We used the algorithm `prune.tree` in S-Plus (Venables and Ripley, 1997) to calculate the tree with optimal number of partitions.

We partitioned the data set into 10 random groups and we fitted ten trees, each one using 9 out of the 10 groups. The tree deviance for the remaining group was estimated. Averaged deviance was calculated across the ten trees. This whole process was repeated 100 times using random partitions of the data set. Tree size ( $n$ ) was selected as the number of partitions that minimize the average cross-validated deviance, and the original tree was pruned to size  $n$ .

The variables selected by CART were used to construct a mixed linear model. The significance of each term was tested using analysis of variance (ANOVA) and alternative models were compared using the Akaike Information Criterion (Akaike, 1974). Variables and interactions not included in the first fitted model were included using a step-wise procedure. Results were grouped according to fixed effects, and significance between groups was estimated by Fisher least significant difference test (LSD) (Hochberg and Tamhane, 1987).

## Results

The *E* and *dt* loci affected all phases during reproductive development and they also affected pod number. Pod number was analyzed as the result of two processes: pod addition duration and rate of pod addition. The results are presented first for pod addition duration followed by the other phases and their relationship to pod addition duration. The combination of statistical techniques used to analyze the data is illustrated when

presenting the results for pod addition duration. Only results from parametric analyses are presented in subsequent sections. Within each section, main effects are presented first, followed by interactions, redundancy between loci and particular cases.

### **Pod Addition Duration**

CART analysis suggested that planting date, growth habit and the loci *E3* and *E5* mediate pod addition duration. Figure 2-4 shows the tree that minimizes the cross-validated deviance (Fig.2-4b), which is achieved after five partitions. The tree shows that planting date is the major environmental variable regulating phase development. The second partition in the hierarchy shows that growth habit controls development, being delayed by *Dt*. Indeterminate soybeans respond differently to planting date depending on the presence of *E5* and then *E3*. For a given growth habit, the loci *E3* and *E5* are not present in both branches of the tree, suggesting interactions between growth habit and *E* loci. A mixed linear model demonstrates that planting date, growth habit, the loci *E3* and *E5* regulate pod addition duration (Table 2-2). In addition, the mixed linear model showed that pod addition duration was also under the control of *E1* and *E2*, and that the responses to planting dates varied between genetic backgrounds. Interactions between planting date, growth habit and between *E3* and *E5* were significant as inferred from the tree structure (Table 2-2).

The duration of the pod addition period varied between planting dates. In early plantings, all NILs had longer phase durations than in late plantings. Because of interactions between planting date and loci, but not between year and loci (Table 2-2), the response to planting date can be attributed mainly to changes in photoperiod. The differences between years in late plantings detected using CART were not significant when we tested them using mixed linear models (Table 2-2). However, the slightly lower

temperatures during 2001 relative to 2002 after R1 in late plantings (Fig.2-2) can explain the longer duration of pod addition observed in 2001.

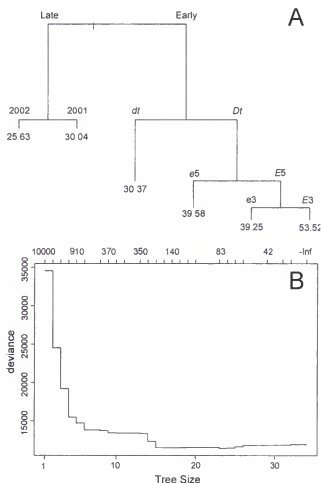


Figure 2-4. Regression tree describing the data structure for duration of pod addition. (A) Minimum spanning regression tree pruned to size 6. Average pod addition durations (days) for each leaf are shown, (B) Cross-validated deviance as a function of tree size.

In the presence of the photoperiod stimuli during the early plantings, the *E* loci in interacting with growth habit regulated the duration of pod addition. While determinate NILs (*dt*) completed setting pods in 30 days on average (Fig.2-4a), NILs carrying the *Dt* allele showed a wide range of variation; from 39 to 53 days for NIL carrying *e* and *E* alleles, respectively (Fig.2-4a). Under long photoperiod, genetic background effects were observed. Harosoy had longer phase duration than Clark NILs (Table 2-3). In Clark

genetic background and determinate growth habit (*dt*), the *E1* and *E2* alleles extended the pod addition duration, being mainly controlled by *E2*. In contrast, *E3* and *E5* had main control of the phase duration in indeterminate NILs (Table 2-3).

Table 2-2. Analysis of variance for factors affecting soybean reproductive development, node number, and rate and duration of pod addition. Terms not included in the table are either not significant or cannot be estimated (e.g., higher order interactions between loci).

Factor	Durations				Pod number per day of pod addition
	R1-OPA	OPA-R5	R5-R7	Pod addition	
PDAT	**	**	**	**	**
BCK	**	NS		**	**
<i>Dt</i>	NS	**	**	**	*
<i>E1</i>	**	**		**	**
<i>E2</i>	**	**		**	NS
<i>E3</i>	NS	**		**	NS
<i>E4</i>	NS	NS		NS	NS
<i>E5</i>	**	**	**	**	NS
<i>E7</i>	NS	NS		NS	NS
<i>E1</i> x <i>E2</i>	**	NS		**	NS
<i>E1</i> x <i>E5</i>	*	NS		NS	NS
<i>E2</i> x <i>E3</i>	NS	*		*	NS
<i>E3</i> x <i>E5</i>	NS	**		*	NS
<i>dt</i> x <i>E1</i>	NS	NS		NS	*
<i>dt</i> x <i>E2</i>	NS	*		NS	NS
PDAT x <i>E1</i>	*	NS		**	NS
PDAT x <i>E2</i>	**	**		**	NS
PDAT x <i>E3</i>	NS	**		**	NS
PDAT x <i>E5</i>	**	**		**	NS
PDAT x <i>dt</i>	NS	**		**	NS

NS: not significant

\* significant 5% \*\* significant 1%

PDAT: planting date and photoperiodic stimuli

BCK: genetic background

These results show that *E* loci interact and have redundant functions in the control of pod addition duration. Strong interactions were evident between *E3* and either *E2* or *E5*, and between *E5* and *E1* (Table 2-2, Table 2-3). Interactions between *E1* and *E3* were only observed in Harosoy NILs (Table 2-3). Our results suggest that *E2* and *E5* have

redundant function. Under long photoperiod during early plantings, there were no differences between genotype *e1E2e3e5* and *e1e2e3E5* in pod addition duration. Furthermore, in the presence of *E3*, the presence of either *E2* (*e1E2E3e5*) or *E5* (*e1e2E3E5*) increased phase duration by approximately 10 days relative to the genotype *e1e2E3e5* (Table 2-3). In Clark but not in Harosoy background, *E1* and *E2* showed redundant functions. In the presence of *E3* and *E5*, pod addition duration remained constant whether *E1* was replaced by *E2* or if both loci were present.

Table 2-3. *E* loci, growth habit and genetic background effects on pod addition duration (days) on different planting dates

Genotype	Early planting				Late planting			
	Clark		Harosoy		Clark		Harosoy	
	<i>Dt</i>	<i>dt</i>	<i>Dt</i>	<i>dt</i>	<i>Dt</i>	<i>dt</i>	<i>Dt</i>	<i>dt</i>
<i>e1e2e3e5</i>	34.8	24.0	35.2	26.1	29.5	24.0	27.1	25.6
<i>E1e2e3e5</i>	37.5	29.5	40.4	32.3	27.8	25.8	28.4	26.8
<i>e1E2e3e5</i>	40.3	34.0	40.2		30.5	26.0	25.4	
<i>e1e2E3e5</i>	35.4	26.0	40	28.3	29.3	22.5	25.5	27.5
<i>e1e2e3E5</i>	39.8		38.7		31.3		32.8	
<i>E1E2e3e5</i>	34.3	33.5			31.0	22.8		
<i>E1e2E3e5</i>	39.5	32.0	49.2	34.5	31.7	21.8	31.8	29.0
<i>e1E2E3e5</i>	45.5	34.3	46.5		31.5	27.0	26.3	
<i>e1e2E3E5</i>	44.5		48		28.3		31.3	
<i>E1E2E3e5</i>	43	37.8	45		30.8	24.5	25.3	
<i>E1e2E3E5</i>	54		55		32.5		28.3	
<i>e1E2E3E5</i>	54.5		64		34.8		30.5	
<i>E1E2E3E5</i>	55				32.3			
Mean	42.9	31.4	45.7	30.3	30.9	24.3	28.4	27.2
LSD (0.05)	7.6	5.7	7.1	4.4	6.4	6.4	6.3	5.4
LSD (0.01)	10.9	8.4	10.3	6.4	9.2	9.3	9.0	7.8

The effects of *E3* on pod addition duration shown in this paper are consistent with previous reports on the effects of these loci on the duration of the flowering period in indeterminate Clark (Summerfield et al. 1998; Asumadu et al., 1998). However, our findings do not support the notion that a longer flowering duration in genotypes carrying *E1* and *E2* propagates into longer pod addition duration. *E1* and *E2* were associated with

delays in the onset of pod addition (Table 2-2; Fig.2-5), delay which can explain the effects of *E1* and *E2* on flowering duration but not on pod addition duration. In contrast, our results indicate that *E1* and *E2* indeed extend pod addition duration in determinate soybeans. In these genotypes, *dt* minimizes the time to onset of pod addition and the potential extension of flowering duration (as inferred from their effects on indeterminate background; Summerfield et al., 1998) would translate into longer pod addition duration.

### **Duration of Time Between Flowering and First Pod (R1-OPA)**

The time between first flower and beginning of pod addition varied between planting dates but not between years. The *E1*, *E2* and *E5* alleles significantly affected the onset of pod addition in interaction with planting date (Table 2-2; Fig.2-5). Genetic background effects were significant (Table 2-2) but of reduced magnitude; Clark NILs set pods  $0.4 \pm 0.13$  days later than Harosoy NILs.

In the absence of dominant *E* loci (genotype *e2e3e5*), the duration of this phase was about five days (Fig. 2-5), which is roughly the period required for embryo growth before cotyledon initiation and for initiated pods to reach 5 mm (Carlson, 1973). This result is consistent with previous observations (Johnson et al., 1960). In early plantings the effects of *E* loci increased the phase duration to 10 days, doubling the period required for pod growth in late planting dates (Fig.2-5). The *E5* allele showed the largest effects on the time from R1 to OPA when interacting with *E1* and *E2* (Fig.2- 5). Variations of this magnitude and larger were observed under extended photoperiod and suboptimal temperature (van Schaik and Probst, 1958; Johnson et al., 1960). Because temperature regimes were similar between years and planting dates (Fig2-2), the observed delays on the period between flowering and pod set were due to the changes in photoperiod between planting dates. While there were no significant differences between years on the

duration of this period, the effects and interactions between planting dates and *E* loci were highly significant (Table 2-2).

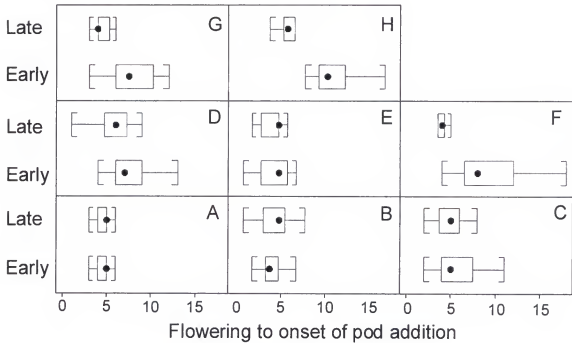


Figure 2-5. Duration between flowering and the onset of pod addition as affected by planting date (early: May planting, late: July-August planting) and *E* loci. A) *e1e2e5*, B) *E1e2e5*, C) *e1E2e5*, D) *E1E2e5*, E) *e1e2E5*, F) *E1e2E5*, G) *e1E2E5*, H) *E1E2E5*. The failure to set pods mediated by the loci *E1*, *E2* and

*E5* in early plantings can be associated with increased flower shedding and embryo abortion in response to longer photoperiods. Under optimal temperature (see Boote et al. (1998) for a detailed response curve of physiological processes to temperature), flower shedding in soybean “Clark” was highly correlated to variations in the time between flowering and the onset of pod addition under different photoperiods (Schaik and Probst, 1958). Fisher (1963) demonstrated that the inability to set fruits by soybeans grown under long photoperiod was associated with male sterility, and that fertility was restored by exposing plants to three consecutive short photoperiods. Recent evidence showed that photoperiod can also be involved in embryo abortion after cotyledon differentiation

(Tischner et al., 2003). Under long photoperiods, flower abortion increased from 49 flowers per plant in the Clark NILs L92-21 (*Dte1e2E3e4e5e7*) to 131 flowers in L74-441 (*DtE1E2e3E4e5E7*) (Wilcox et al. 1995; see Table 2-1 for NIL genotypes). In early plantings, the NIL L74-441 showed a significant 3-day delay between flowering and the onset of pod addition relative to NIL L92-21. Higher pod setting efficiency was reported under short photoperiod (Board and Settini, 1986; Thomas and Raper, 1976). Recent results linked ovule abortion at the time of fertilization with quantitative trait loci for flowering date indicating that these two processes shared a set of genes mediating the responses to photoperiod.

### **Duration between OPA and R5**

The onset of seed growth (R5) is associated with the duration of pod addition (Fig.2-1; Fig.2-3). At the onset of seed growth (R5), assimilates are increasingly directed to the seed, decreasing assimilate availability for setting new pods. At the same time, the addition of new nodes on the main stem extends the duration of pod addition by creating new reproductive sites and source size. This paper tests the hypothesis that *E* loci regulate the duration between onset of pod addition and R5.

Growth habit showed major control on the duration of this phase, overriding the control of *E* loci (Table 2-4). The large effects of *dt* and *e* alleles on this phase duration is not surprising since *dt* inhibits the addition of new nodes (Wilcox et al., 1995).

The *E1*, *E2*, *E3* and *E5* alleles extended the time to R5 on indeterminate background NILs (Table 2-4) but had only minor control on determinate (*dt*) NILs. The effects of *E2* and *E3* on phase duration are consistent with previous reports on the lengthening of the period R2 to R5 (Guffy et al., 1991). The strong interaction between these loci with planting date (Table 2-2) and the lack of interaction with year ( $P=0.6394$ )

indicate that the regulation of this phase is triggered by photoperiod and mediated by the *E* and *dt* loci.

The *E* loci interacted with each other and had redundant functions. *E3* interacted with *E2* and *E5* (Table 2-2) extending the time between the onset of pod addition and R5 from 23 to 32 days when all three loci were present (Table 2-4). *E1* did not interact with *E2* and *E3* (Table 2-2, lines 10-11) as was the case for the phase R1-OPA (Table 2-2) and duration of flowering (Summerfield et al., 1998; Asamadu et al., 1998).

Table 2-4. *E* loci and *dt* control of the duration (days) between the onset of pod addition and R5 (beginning seed growth) on different planting dates

Genotype	OPA-R5			
	Early planting		Late planting	
	<i>Dt</i>	<i>dt</i>	<i>Dt</i>	<i>dt</i>
<i>e1e2e3e5</i>	19.3	13.6	15.3	9.0
<i>E1e2e3e5</i>	22.6	11.9	16.7	10.2
<i>e1E2e3e5</i>	23.4	9.8	15.4	10.3
<i>e1e2E3e5</i>	21.4	13	16.9	8.5
<i>e1e2e3E5</i>	20.8		19.2	
<i>E1E2e3e5</i>	23.3	12.3	19.5	11.3
<i>E1e2E3e5</i>	22.0	12.6	17.5	10.1
<i>e1E2E3e5</i>	31.6	12.5	18.2	9.8
<i>e1e2E3E5</i>	32.6		17.8	
<i>E1E2E3e5</i>	30.0	16	17.3	12.3
<i>E1e2E3E5</i>	36.1		18.8	
<i>e1E2E3E5</i>	36.9		18.9	
<i>E1E2E3E5</i>	39.5		19.3	
LSD (0.05)	5.4	2.9	3.9	1.9
LSD (0.01)	7.8	4.1	5.5	2.7

The loci *E2* and *E5* seemed to have redundant function. Replacing *E2* for *E5* in an *E3* background did not change the phase duration (compare lines 8 and 9 relative to line 4 in Table 2-4). The presence of both *E2* and *E5* in an *E3* background has similar effects on this phase duration as the individual loci (Table 2-4).

A longer time to sink development would increase the pod addition period due to low competition for assimilates in the absence of rapidly growing seeds. The mechanism of this hypothesis is the conceptual framework illustrated in Fig.2-1, is that the longer the time between the onset of pod addition and R5, the longer the period of pod addition. Fig.2-6 shows that the two periods were strongly correlated despite differences in planting date and growth habit. Indeterminate soybeans had longer pod addition duration as the time to R5 was delayed. In contrast, R5 in determinate soybeans occurred earlier relative to indeterminate soybeans, establishing a strong sink early, which reduces the assimilate pool for pod setting. Therefore, NILs with the *dt* allele had a shorter pod addition duration.

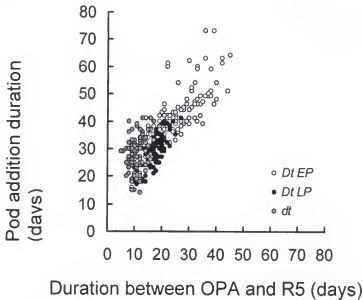


Figure 2-6. Relationship between pod addition duration and the onset of seed growth (as measured by time between OPA to R5). Early (EP) and Late (LP) denote time of planting. *Dt* and *dt* are dominant and recessive alleles for growth habit.  $y = 15.5 (\pm 0.76) + 0.97 (\pm 0.04) x$ ;  $df = 359$ ;  $R^2 = 0.63$ ;  $P < 0.0001$

Similar to the patterns observed for determinate NILs, late planted NILs had shorter time to R5 and shorter pod addition duration. The results provide genetic evidence supporting the model presented in Fig.2-1 and the hypothesis that *Dt* and *E* loci regulate,

at least partially, the duration of pod addition by determining the time to R5. Recent evidence linked pod abortion to QTLs regulating time to flowering and water use efficiency (Tischner et al., 2003). This result suggests that both photoperiod and carbon assimilation are involved in the process of pod abortion. Similarly, Egli and Bruening (2000) showed that both the availability of photoassimilates and the duration between flowering and the onset of seed growth modulates seed number.

### **Final Pod Number and Pod Addition Duration**

The ratio between pod number and pod addition duration varied with planting date, genetic background, growth habit and allele *E1* (Table 2-1; Fig.2-6). Late plantings, Harosoy NILs, determinate growth habit and *e1* allele decreased the duration of pod addition. Similar effects of planting date were observed in commercial varieties (Kantolic and Slafer, 2001). Shorter photoperiod during the growing period of late plantings shortened pod addition duration (Table 2-3) and decreased the number of branches (data not shown), which in turn reduced the potential sites for addition of new pods.

Due to the earlier termination of node differentiation in the main stem in determinate soybeans relative to indeterminate ones, the addition of new pods relies heavily on the appearance of new branches. Addition of new pods in branches requires a lag time to grow the vegetative structures before the onset of pod addition in new nodes. However, the addition of new branches allows the simultaneous addition of pods. The rate of pod addition was related to the number of nodes on branches (*NNB*) between zero and twenty fives nodes [ $RPA = 1.02 (\pm 0.04) + 0.059 (\pm 0.005) \bullet NNB$ ;  $df=345$ ;  $R^2=0.27$ ], above which the rate of pod addition reached a maximum value. The earlier change of the apex from vegetative to reproductive in determinate NILs increases assimilate

partitioning to reproductive structures but also branch nodes, which can explain their relative higher rates of pod addition (Fig.2-7)

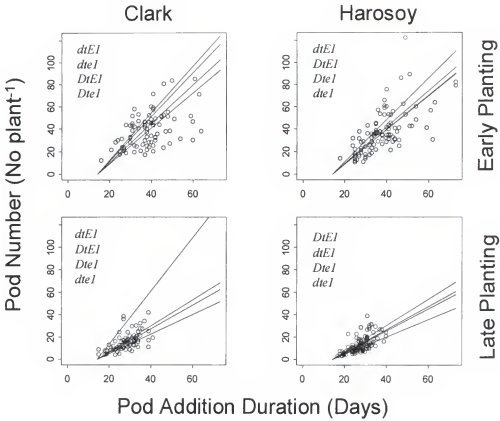


Figure 2-7. Final pod number as a function of pod addition duration, genetic background, planting date, growth habit, and locus *E1*. Genotypes indicated in the legend are arranged from highest to lowest slope.

The allele *E1* was associated with increased radiation use efficiency in a study reported by Ellis et al. (1998). Higher availability of assimilates could increase the rate of pod addition by limiting the abortion of small pods (< 5mm). In addition, NILs with the *E1* allele had a significant higher number of branches and nodes in branches (data not shown) relative to other NILs, allowing the plant to set pods simultaneously on a larger number of sites relative to the NILs carrying the recessive allele.

A recent study suggests that *E1* may encode phytochrome B (Tasma and Schoemaker, 2003). Near isogenic lines carrying the recessive allele *e1* would be

impaired in the perception, transduction and elicitation of biological responses associated with red light. Heindl and Brun (1983) demonstrated that pod number increased with red light enrichment and that this effect was not due to increases in photosynthesis. The reduction in the rate of pod addition in genotypes carrying the *e1* allele may be related to the inability to perceive light signals.

### **Duration between R5 and R7**

The duration between R5 and R7 is considered a good estimator of seed fill duration (Nelson, 1986). Seed fill duration as defined above varied with planting date and growth habit (Table 2-2, 2-5). Harosoy NILs showed trends toward longer seed fill duration than Clark. Seeds for NILs with genotype *e1e2e3* were shown to be heavier in Harosoy than Clark (Curtis et al., 2000; Wilcox et al., 1995) requiring longer seed fill duration to reach maximum seed size. Seed fill duration was shorter in late than early plantings (Table 2-5) in response to differences in photoperiod between planting dates. Under short photoperiods during the late plantings, assimilate partitioning to reproductive sinks increased (Thomas and Raper, 1976; Cure et al., 1982) and pod load decreased in response to shorter pod addition duration (Table 2-3). The increased assimilate availability per pod increased seed growth rate and decreased seed fill duration (Swank et al., 1987). Previous studies showed that seed weight in NILs carrying recessive *e* alleles was higher relative to those carrying the dominant ones (Wilcox et al., 1995; Guffy et al., 1991; Curtis et al., 2000). Because photoperiod-insensitive lines showed lower seed filling duration, these NILs must have higher seed growth rates. Also, *E* allele-carrying NILs have higher sensitivity to photoperiod that slows down individual seed growth rate, thus can add more pods extending pod addition and seed fill duration (Wilcox et al., 1995; Guffy et al., 1991).

The loci *dt* increased seed fill duration relative to indeterminate NILs. Determinate soybean increased assimilate partitioning to reproductive organs relative to leaves and stems, decreased pod addition duration (Table 2-3) and consequently pod number (Fig.2-6), and therefore increased the availability of assimilates per pod. Despite the higher expected seed growth rate, determinate soybeans showed longer seed fill duration as shown in previous experiments (Guffy et al., 1991). These variations may be related to the differences between NILs on the position of the pods at the onset of pod addition. While determinate soybeans begin pod setting almost simultaneously at all positions in the canopy, indeterminate lines add the last pods in the upper nodes. Early set pods have a higher ratio between pod wall to seed mass than later added pods suggesting that a smaller seed size is attained in later added pods (Fraser et al., 1982).

Table 2-5. Effects of *E* loci and *Dt* on seed fill duration for different years, planting dates and genetic background

Genotype	2001		2002	
	Early	Late	Early	Late
' <i>dte5</i>	30.8	30.6	31.2	26.8
' <i>Dte5</i>	25.8	21.2	26.0	22.9
' <i>DtE5</i>	24.5	21.4	35.4	23.2
LSD (5%)	7.0	4.4	8.2	4.7
LSD(1%)	9.9	6.2	11.7	6.4

### Discussion and Conclusions

Understanding soybean development and its genetic control during the reproductive stages is of importance for physiologists, modelers and plant breeders who are interested in predicting and increasing soybean yields. This chapter showed that the *dt* locus and *E* loci regulate seed-filling duration, and the photoperiodic response of the duration from first flower to first pod (Fig.2-5), the duration of the critical period of pod addition (Table 2-3, Fig.2-4) and the time to R5 (Table 2-4). These results suggest two mechanisms for

the regulation of pod addition duration. One is associated with early embryo abortion (or failure in ovule fertilization), which is expressed as delays in the onset of pod addition caused by marginal daylength. The other, consistent with the hypothesis underlying the model of Figure 2-1, relates the termination of the pod addition period *via* pod abortion due to a reduced availability of assimilates. *E* loci regulate pod addition duration through the determination of the onset of seed development (R5) as shown by the association between pod addition duration and time to sink development (Fig.2-6). Finally, *E* and *dt* loci controlled fruit number by regulating pod addition duration.

Intercepted radiation during the period of fruit addition exerts strong control on fruit number and yield (Kantolic and Slafer, 2001). The duration of this critical period was shown to be under photoperiodic control. This paper provides evidence for the role of the dominant alleles at the *E1*, *E2*, *E3* and *E5* loci, on lengthening the pod addition duration phase in response to photoperiod. The action of these loci showed redundant action, which was related to genetic background. The extension of pod addition duration under long photoperiods increased pod number (Fig.2-6) presumably as the result of increases in intercepted radiation and increased photoassimilates. However, in the case of the *E1* locus, pod number increased also as a consequence of a higher rate of pod addition, probably associated with increased radiation use efficiency as reported by Ellis et al. (1998). Other research showed contrasting results, such that *E1* decreases seed number (Guffy et al., 1991; Curtis et al., 2000). This contrasting result may arise from genotype by environment interactions. The locus *E1* has the largest effects on time to flowering and onset of pod addition (Fig.2-4; Curtis et al., 2000). Under longer photoperiods in Illinois and Ontario than in Florida, the locus *E1* could have delayed

flowering and the onset of pod addition, thus decreasing the duration of the reproductive period, pod addition duration and reproductive efficiency, which can explain the reduction of seed number.

Because of the association between pod addition duration and pod number, increasing this phase duration can increase soybean yields. As this phase is under photoperiod control, increasing the photoperiod sensitivity can extend pod addition duration. However, this approach has the limitation that a concurrent extension of pod addition duration can lead to a reduction in reproductive efficiency because the latter is also under photoperiodic control (Zhang et al., 2001; Thomas and Raper, 1984; Board and Settini, 1988; Fisher, 1963). In other words, photoperiods marginally too long would reduce pod number because of flower abortion.

An alternative approach to increasing pod number is to accelerate the onset of pod addition. This implies reducing photoperiod sensitivity before but not after the onset of pod addition. Modeling studies assumed that the durations from flowering to onset of pod addition, and from flowering to onset of seed addition are proportional (Mavromatis et al., 2001). The co-regulation of these phase durations would hamper the implementation of this type of strategy. The present study, however, shows that this proportionality only holds for NILs carrying loci *E1* and *E2*, which delays the onset of pod addition more than the onset of seed growth. In contrast, *E3* mediates the photoperiodic response after, but not before, the onset of pod addition (Table 2-2; Fig. 2-5). Therefore, *E3* in combination with other loci has the potential to extend pod addition duration in improved varieties as shown for Clark and Harosoy NILs (Table 2-3; Fig. 2-4). Accelerating the onset of pod addition can increase harvest index and yield by reducing flower abortion prior to the

onset of pod addition, increasing reproductive efficiency and, reallocating assimilates from vegetative sinks to set and fill reproductive organs. Early onset of fruit addition was related to increases in yields of peanut (Gifford et al., 1984) and soybean (Boote et al., 2001).

A third strategy to increase fruit number in soybean would consist of increasing the relative duration of the reproductive period with respect to the vegetative period (Kantolic and Slafer, 2001). This strategy requires at least semi-independent photoperiodic regulation of the pre and post-flowering periods. This study suggests that *E5* may be a good candidate to achieve this goal since it has minor, if any effects on flowering time (data not shown) but it plays a major role in regulating pod addition duration. Indeed, the combination of *E3* and *E5* showed the largest pod number of all the isolines used in this study (Fig.2-4).

Seed fill duration was shown to be correlated with seed yield (Smith and Nelson, 1986; Boerma and Ashley, 1988; Guffy et al., 1991) and has been used as a criterion for selection in plant breeding for higher yield (Nelson, 1986). Minor effects of *E* loci were shown associated with the regulation of the period between R5 and R7 in the experiments conducted in Florida (Table 2-5); generally in the form of shortening of the period when *E5* allele was present, similar to results reported by Guffy et al. (1991). These results contrast with those in other experiments conducted at higher latitudes in which seed fill duration increased with dominant *E* alleles (Wilcox et al., 1995; Curtis et al., 2000).

Because *E* loci also regulate both time to flowering and maturity (e.g., Cober et al., 1996), *E* loci effects on seed fill duration may be partially confounded with environmental effects. This can be illustrated from Wilcox et al. (1995) experiments in

which *E1E2E3* genotypes flower 30 days later than *e1e2e3* NILs. Lower temperatures with late flowering could have increased seed fill duration (Wilcox et al., 1995). The extension of seed filling period with longer photoperiods and dominant *E* alleles (Wilcox et al., 1995), can also be the consequence of a longer pod addition duration, an increased duration of the individual seed filling period, or both (Fig.2-3). From Guffy et al. (1991) there is little evidence for *E* loci affecting individual seed filling duration. However, strong relationships were shown between seed number and yield (Guffy et al., 1991; Shibles et al., 1975; Egli, 1998). A reanalysis of Curtis et al. (2000) data shows that yield (*Y*) differences between NILs are related to variations in seed number (*SN*) [ $Y = 0.0018 + (0.0002) \bullet SN$  ( $P < 0.0001$ ;  $R^2 = 0.896$ )], but not to seed size ( $P = 0.26$ ). This result suggests that at least part of the effects of the *E* loci on seed fill duration is by mediating the photoperiodic effects on pod addition or single seed growth demand as shown in this study (Table 2-3; Fig.2-4). Future research is needed to further understand the effects of *dt* and *E* loci on seed fill duration.

Kantolic and Slafer (2001) proposed to increase the duration between R3 and R6 at the expense of shortening the duration of the vegetative phase to increase soybean yields. This assumes the absence of yield component compensation. Results presented here support this strategy, suggesting that *E3* and *E5* loci can be used to implement this genetic improvement strategy. However, the selection strategy can be improved by using pod addition duration instead of the duration between R3 to R6, and by including the duration between flowering and onset of pod addition as additional selection criteria.

We studied the genetic control of response to photoperiod mediated by *dt* and *E* loci during the reproductive period, and to evaluate their effects on fruit number.

Previous research described in the literature reported the effects of *E* loci on time to flowering and maturity. However, we had incomplete knowledge about the effects of *E* loci on critical sub-periods of the reproductive development of soybean. A field experiment that exploits variations in photoperiod by changing planting date was conducted in two years to test the hypotheses:

- The *dt* and *E* loci regulate the duration of the following periods: a) from first flower to first pod; b) pod addition; c) seed filling; and d) from first flower to the onset of seed development.
- *E* loci regulate pod number by affecting the rate of pod addition.
- *E* loci regulate duration of pod addition by regulating the onset of seed development.

This Chapter showed that the *dt* locus and *E* loci regulate the photoperiodic response of the duration from first flower to first pod (Table 2-2, Fig.2-5), the duration of the critical period of pod addition (Table 2-2, Table 2-3, Fig.2-4), time from OPA to R5 (onset of seed filling) (Table 2-4) and the seed-filling duration as estimated by the time between R5 and R7 (Table 2-5).

We showed that *E* loci regulate pod addition duration through the determination of the onset of seed development as shown by the association between pod addition duration and time between OPA and R5 (Fig.2-6). Finally, *E* and *dt* loci controlled fruit number by regulating pod addition duration. The results obtained do not support conclusively the relationship between rate of pod addition and pod number.

Finally, methods and approaches need to be developed to design ideotypes based on genetic information. Simulation frameworks based on crop models that incorporate genomic information (White and Hoogenboom, 1996) and optimization algorithms can

provide the basis for designing ideotypes for target environments and test selection strategies as discussed above.

## CHAPTER 3

### A GENE-BASED APPROACH TO SIMULATE SOYBEAN DEVELOPMENT AND YIELD RESPONSES TO THE ENVIRONMENT

#### **Introduction**

With the world's population increasing and global grain demand projected to double by the middle of this century, new ways of increasing yields while preserving natural habitats and diversity must be found (Trewavas, 2002; Tilman et al., 2002). Another Green Revolution is needed in years to come under a scenario of water limitations, less favorable environmental conditions, and exhaustion of past sources of growth (Huang et al., 2002). As in the past, plant breeding, now empowered by molecular techniques, will constitute the backbone for breaking productivity constraints (Huang et al., 2002; Knight, 2003). Recent advances in plant genomics promise to affect many aspects of plant genetic improvement (Somerville and Somerville, 1999). However, the realization of the potential contributions of functional genomics to plant breeding is dependent on the development of a robust genetic engineering discipline, which should provide methods and tools to understand intrinsically complex biological systems and to predict phenotypes such that rational changes can be designed (Somerville and Somerville, 1999; Cooper et al., 2002; Chapman et al., 2003). Furthermore, differences in crop performance and physiology between field versus laboratory conditions are well known, requiring extended field testing to anticipate benefits and unexpected pleiotropic effects in improved and transgene varieties (Strauss, 2003). The development and application of mathematical concepts and systems approaches to uncover principles

underlying biology at molecular, cellular and organism levels is emerging under the name of computational systems biology (Kitano, 2002).

Crop models have the potential to become powerful genetic engineering tools. These dynamic process-oriented models (e.g., DSSAT, Jones et al., 2003; APSIM, Keating et al., 2003; van Ittersum et al., 2003) incorporate the state of the knowledge of environmental and managerial effects on crop growth and development by simulating the effects of climate on physiological processes, soil and nutrient dynamics. Differences between genotypes are taken into account by a set of parameter-controlling morphological and physiological traits named *genetic coefficients* (Hunt and Boote 1998).

The systematic and modular structure of crop models (Jones et al., 2003) provides the means to link and evaluate the effects of manipulations at cellular and molecular levels on the plants at the organism and field scales. This property of crop models can help us understand and evaluate pleiotropic effects of genes, disentangle complex traits and genetic by environment (GxE) interactions, assist multigene engineering, and reduce and guide field-testing. The vast majority of agronomic traits are quantitative in nature and polygenetically controlled (Daniell and Dhingra, 2002; Stuber et al., 2003) leading to strong GxE interactions (Allard and Bradshaw, 1964) and gene-gene interactions (Lark et al., 1995; Lark et al., 1994; Orf et al., 1999a).

Genotypic differences in current crop models are paradoxically phenotypic in nature, thus limiting their applicability. Further limitations arise from the fact that genetic coefficients are seldom measured; instead, numerical optimization algorithms that require intensive computation and large data sets are used (Hunt et al., 1993; Mavromatis et al.,

2001; Grimm et al., 1993). The first conceptual attempt to overcome the problem was published by White and Hoogenboom (1996) in Genegro, a process-oriented model that incorporated effects of seven genes affecting phenology, growth habit and seed size. Genetic coefficients in Genegro are estimated from genes and a set of linear functions. Genegro accurately predicted dry bean (*Phaseolus vulgaris* L.) development but poorly explained yield variations between sites (Hoogenboom and White, 1997). Recently, Hoogenboom and White (2003) modified Genegro to account for the effects of temperature on photoperiod sensitivity regulated by the gene *Tip*, which improved prediction skill. Similar modeling approaches were used to incorporate quantitative trait loci (QTL) effects on leaf elongation rate (Reymond et al., 2003, Tardieu, 2003), specific leaf area (Yin et al., 1999), plant height, pre-flowering duration, carbon partitioning to spike, spike number, and radiation use efficiency (Yin et al., 2003) in barley (*Hordeum vulgare* L.), and the time to flowering (Stewart et al., 2003; Cober et al., 2001; Upadhyay et al., 1994a), and flowering duration (Summerfield et al., 1998) in soybean [*Glycine max* (L.) Merrill].

There are no process-oriented models that incorporate gene actions for soybean, despite previous photothermal models that predict time to flowering and flowering duration (Stewart et al., 2003; Upadhyay et al., 1994a; Summerfield et al., 1998; Cober et al., 2001) based on the genetic makeup of *E* loci of soybean near-isogenic lines (NILs). Neither former gene-based approaches to simulate soybean time to flowering, nor gene-based models for any other crops were validated or tested for their ability to predict plant growth and development after genotyping commercial cultivars.

Previous research provided experimental evidence of *E* loci control of reproductive development (Wilcox et al, 1995; Curtis et al., 2000; McBlain et al., 1987); in particular, the effects of photoperiod on the onset of pod addition, pod addition duration, and the onset of seed growth in soybean (Chapter 2). Based on these advances, a gene-based model for soybean was developed by incorporating gene action into the CROPGRO-Soybean model (Boote et al., 1998). The model was evaluated using an independent data set and evaluated to predict time to maturity of public soybean cultivars grown in variety trials.

### **Materials and Methods**

The prediction of crop development is critical for simulation of plant growth. The generation of leaf area, assimilate partitioning, the duration of critical events, and the timing of responses to environmental stresses are under the genetic control of developmental processes. Soybean development in CROPGRO is roughly subdivided into the time from emergence to flowering, from flowering to the onset of a) pod addition, and b) seed addition, and from the onset of seed addition (growth) to physiological maturity. The duration between emergence and flowering is further subdivided into three phases: a) juvenile phase, b) an inductive phase, and c) a phase that starts at flower initiation and ends when the first flower becomes visible. The duration of the vegetative phase varies between determinate and indeterminate soybeans, and this difference is coded by the time to flowering plus the genetic coefficient FL-VS, which determines the physiological time between flowering and the end of differentiation of nodes in the main stem. FL-VS generally coincides with R5. In both growth habit types, the end of leaf area expansion, which includes leaf area in branches, ceases at a time near the end of pod addition.

### Simulation of Soybean Development and Pod Addition

CROPGRO uses a multiplicative function of photoperiod ( $P$ ) and temperature ( $T$ ) to model developmental progress during different growth phases (Grimm et al., 1994; Grimm et al., 1993; Jones et al., 1991).

$$R(t) = f(P) \times f(T) \quad (1)$$

The  $R(t)$  model predicts relative development rate with maximum rate standardized to 1.0. At optimum temperature and photoperiod, the rate of progress in physiological days equals the rate of progress in calendar days. When conditions deviate from the optimum, the rate of development per day decreases, becoming a fraction of a physiological day. The multiplicative model holds for a period beginning after the juvenile phase. During the juvenile phase, the plant is not receptive or sensitive to changes in daylength, and development is only a function of temperature. Each phase has its own “developmental accumulator” starting at a unique point in time, and when it reaches a threshold an event is triggered and a phase finishes. This non-linear model (Fig. 3-1A) has been shown to have high predictive capabilities (Grimm et al., 1993; Grimm et al., 1994; Mavromatis et al., 2001; Mavromatis et al., 2002). Alternative approaches to predict time to flowering have used multiple regression models (Hadley et al., 1984; Summerfield et al., 1993), genetic algorithms (Pabico et al., 1999), neural networks (Elizondo et al., 1994; Welch et al., 2003), and logistic or linear functions (Sinclair et al., 1991).

CROPGRO simulates daily cohorts ( $i$ ) of pods and seeds without distinguishing between locations on branches or main stem. Simulation of pod addition in CROPGRO is based on the most limiting factor between flower production ( $FLWP$ ), the maximum rate

of pod addition (*PODADD*), and today's carbon (C) and nitrogen (N) remaining after seed growth,

$$PODN(i) = \min \{ PODADD(i), FLWP(i), PGLEFT(i) / (SHMAXG \cdot AGRSH) \} \quad (2)$$

where *PODN*(*i*) is the pod number added on day *i*, *PGLEFT*(*i*) is the mass of C available for shell growth after seed growth is accounted for on day *i*, *SHMAXG* is the maximum shell growth rate and *AGRSH* is the C requirement for shell growth.

Calculation of *PODADD* requires the estimation of the maximum load of pods that the canopy can support on day *i* as a function of availability of assimilates at current temperature and irradiance (*PGAVLR*) and the duration of pod addition (*PODUR*). This *PODADD* estimate is affected by environmental conditions as,

$$PODADD = PGAVLR / k \cdot 1 / PODUR \cdot f(T) \cdot f(P)^{1/3} \cdot \min \{f(W), f(N)\} \quad (3)$$

where *k* is a constant that accounts for the carbohydrate cost of pod and seed production, *PODUR* is the photothermal time from first pod added to when the crop reaches maximum pod load, and *f*(*W*) and *f*(*N*) are functions of water and nitrogen availability, respectively. The second term in equation (2), *FLWP*, accounts for the number of flowers that have developed and are ready to form pods (this is not normally limiting in CROPGRO). Temperature and photoperiod can cause flower abortion, thus limiting the addition of new pods. The third term in equation (2) accounts for effects of C limitations on pod setting, and only acts near the end of pod addition when full pod load is present. It is assumed that all of a given day's produced flowers have the potential to be converted into pods if enough C is available to grow the pods for at least one day when the time for pod addition occurs. Water and nitrogen stress will reduce photosynthesis, thus reducing C availability and pod addition for a given day. Also, this term becomes relevant to

simulate the dynamics of pod addition. As new pods and seed are added, the amount of C available to set new pods become scarcer. At full pod load,  $PGLEFT$  becomes equal to zero, defining the end of pod addition.

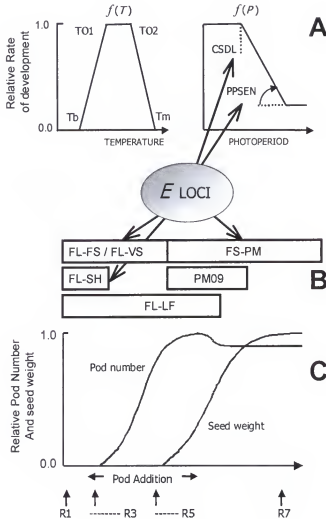


Figure 3-1. Representation of functions used in CROPGRO-Soybean to account for shown photoperiod and temperature effects on soybean developmental phases and their relation to  $E$  loci (A), genetic coefficients for different phase duration (B), and yield components (C).  $T_b$  is the base temperature below which there is no development,  $T_{O1}$  and  $T_{O2}$  define the plateau,  $T_m$  is maximum temperature, CSDL defines the photoperiodic threshold below which relative development is maximum, PPSEN denotes photoperiod sensitivity. See Table 3-2 for other genetic coefficient definitions.

## Data

A set of soybean near-isogenic lines (NIL) (Table 3-1) was used to calibrate a subset of CROPGRO genetic coefficients (Table 3-2). This set of NILs were grown in 2001 and 2002 in early and late planting dates under non limiting conditions of water and nitrogen in Gainesville, FL. Time of first visible flower, onset and end of pod addition, first seed on the upper four nodes (R5), physiological maturity (R7) were recorded at 2-4 day intervals. Pod number were measured at harvest. Experimental details are provided in Chapter 2.

Table 3-1. List of soybean near-isogenic lines used for model development and for molecular markers evaluation

Growth habit & <i>E</i> loci	Name	Molecular marker evaluation
<i>Dt e1 e2 e3 E4 e5 E7</i>	L71-920	Yes
<i>Dt e1 e2 e3 E4 E5 E7</i>	L97-2076	
<i>Dt e1 e2 E3 e4 e5 E7</i>	L92-21	Yes
<i>Dt e1 e2 E3 E4 e5 E7</i>	L63-3117	Yes
<i>Dt e1 e2 E3 E4 E5 E7</i>	L94-1110	
<i>Dt e1 E2 e3 E4 e5 E7</i>	L63-2404	
<i>Dt e1 E2 E3 E4 e5 E7</i>	Clark	Yes
<i>Dt e1 E2 E3 E4 E5 E7</i>	L92-1195	
<i>Dt E1 e2 e3 E4 e5 E7</i>	L80-5914	
<i>Dt E1 e2 E3 E4 e5 E7</i>	L66-432	
<i>Dt E1 e2 E3 E4 E5 E7</i>	L97-4081	Yes
<i>Dt E1 E2 e3 E4 e5 E7</i>	L74-441	Yes
<i>Dt E1 E2 E3 E4 e5 E7</i>	L65-3366	Yes
<i>Dt E1 E2 E3 E4 E5 E7</i>	L98-2064	Yes
<i>dt e1 e2 e3 E4 e5 E7</i>	L80-5882	

Soil parameters for the soil Millhopper Fine Sand (loamy, silic, hyperthermic paleudults) are from DSSAT (Jones et al., 2003). Daily weather data were measured using an automated weather station ~50 m from the experimental plots (temperature, rainfall, solar radiation). The data are available at (<http://plaza.ufl.edu/theagguyl/>).

### Parameterization of CROPGRO-Soybean to Include Genetic Information

CROPGRO was linked to NILs genetic makeup in a two step process. First we estimated a set of genetic coefficients, which are involved in characterizing soybean responses to photoperiod and temperature. Second, the estimated genetic coefficients were modeled as a function of *E* loci using multiple linear regression.

A systematic approach was used for the estimation of the genetic coefficients (Hunt and Boote, 1998) listed in Table 3-2. The approach used as first guesses for the genetic coefficients the set for maturity group 0 in DSSAT (Jones et al., 2003). These coefficients (Table 3-2) were modified in sequence using selected isolines and planting dates. Final values are available with the gene-based model upon request to the author.

Soybean near-isogenic lines carrying only recessive alleles are impaired in the perception or transduction of the photoperiodic signal. These loss-of-function lines grown under short photoperiod ensures the absence of photoperiodic effects on plant development, allowing us to estimate the thermal components of the photothermal time with minimum influence of photoperiod. The lines L71-920 and L80-5882 grown in late plantings (short photoperiod) best satisfy these conditions, and they were used to estimate the thermal component of the photothermal time between emergence to flowering and from flowering to the onset of pod addition. Because L71-920 has indeterminate growth habit, only L80-5882 was used to estimate the photothermal time between flowering and first seed and from first seed to physiological maturity. Both lines grown in late plantings were used to estimate the thermal component of the photothermal time between flowering and the end of canopy expansion (FL-LF), and the proportion of time between first seed and physiological maturity that the last seed is normally formed (PM09). Because no data were available for an accurate calibration, the end of pod addition was used as a proxy

variable for setting the upper bound for FL-LF, and lower bound for PM09. Modification of the genetic coefficient PM09 was subsequently necessary to prevent the underestimation of the pod addition duration.

Table 3-2. Selected genetic coefficients in CROPGRO-Soybean controlling plant development in soybean, potential associations with *E* and *dt* loci, and variables used for parameter estimation

Genetic coefficient	Initial Value	Plant trait	Potential loci effects	Variables used for parameter estimation†	Units
PPSEN	0.171	Photoperiod sensitivity	<i>E</i> and <i>Dt</i>	R1	hr <sup>-1</sup>
CSDL	14.1	Critical photoperiod	<i>E</i>	R1	hr
R1PPO	0.189	Reduction in CSDL after R1	<i>E</i>	R7	hr
EM-FL	16.8	Emergence to flowering	<i>E</i>	R1	PTD*
V1-JU	0.0	Juvenile phase	<i>E</i> 1	R1	TD*
FL-FS	13.0	Flowering to first seed	<i>E</i> and <i>dt</i>	FL-VS & R5	PTD
FL-VS	26.0	First flower to last leaf on main stem	<i>E</i> and <i>dt</i>	R5	PTD
FS-PM	30.0	First seed to physiological maturity	<i>E</i>	R7	PTD
FL-SH	5.0	Flowering to onset of pod addition	<i>E</i>	R1, OPA	PTD
FL-LF	26.0	Flowering to end of leaf area expansion	?	R1, EPA	PTD
PM09	0.35	Proportion of time between first seed and physiological maturity that the last seed is normally added	?	R1, EPA	---

\*PTD denotes photothermal days, TD denotes thermal days.

† R-Stages and Abbreviations: R1 (first flower), R5 (presence of a seed greater than 3mm in a pod in the upper four nodes), OPA: onset of pod addition (when 50% of the plants have a pod greater or equal to 5 mm anywhere on the plant), EPA: end of pod addition, R7 (pod changing color anywhere on the plant).

Gain of function near-isogenic lines carrying the *E* alleles grown in early plantings will perceive and transduce photoperiodic signal. All lines but L71-920 and L80-5882 were used to estimate the critical photoperiod first, and the photoperiod sensitivity in a second step, to predict time to flowering. Photoperiod sensitivity, however, increases

after flowering (Piper et al., 1996). The coefficient R1PPO accounts for this effect by decreasing the value of CSDL. This coefficient was estimated using time to physiological maturity.

A final step used both plantings and gain of function lines to adjust photothermal times between flowering and last leaf formed in the main stem (FL-VS), between flowering and first seed (FL-SD), and time from first seed to physiological maturity (SD-PM). Photothermal time from flowering to first seed in gain of function lines was estimated as a fraction of the photothermal time FL-VS. We used observations made for the stage R5, which typically coincides with the expansion of the last leaf on the main stem, to estimate FL-VS. The average ratio between FL-SD and FL-VS is 0.56 for several standard cultivars of different maturity groups in CROPGRO-Soybean (Jones et al., 2003). FL-SD was then estimated as  $(FL-VS) \cdot 0.56$ . SD-PM was subsequently estimated using data measured for R7.

Results from reciprocal transplant experiments suggest that the locus *E1* is involved in extending the juvenile phase (Upadhyay et al., 1994b). Near-isogenic lines carrying the *E1* allele were used to estimate the duration of the juvenile phase.

### **Parameter Estimation, Model Verification and Evaluation**

Calibration, evaluation, verification and validation of numerical models have been subject of intensive study (e.g., Oreskes et al., 1994; Kobayashi and Salam, 2000; Pachepsky et al., 1996; Colson et al., 1995b). For the purpose of parameter estimation or calibration and model evaluation, this paper followed the approach taken by Hunt and Boote (1998) and Grimm et al. (1993) among others. Parameters were estimated to minimize the root mean square error (RMSE),

$$RMSE = \sqrt{\frac{1}{n} \sum_n (x_i - y_i)^2} \quad (4)$$

where  $n$  is the number of observations,  $x_i$  and  $y_i$  are predicted and observed values, respectively.

For model evaluation is important to use the measurements not used during the calibration of the model. Measurements for pod addition duration and pod number are completely independent except for the NILs L71-920 and L80-5882; information on the end of pod addition for only these two lines were used for calibration purposes. To make the model evaluation more rigorous, observations taken during the 2002 season were only used for model evaluation.

The model was further evaluated by its capacity to reproduce physiologically robust relationships arising from observed data (see Chapter 2), which can provide confirmation about the underlying hypotheses on which the model is built. The evaluation was performed over processes under direct control of genetic coefficients (e.g., time to flowering and maturity) as well as on the constancy of the relationships between pod addition duration and the time to the onset of seed growth. The approach used in this paper, which attempts to confirm the model instead of validating it, prevents the fallacy of affirming the consequent (Oreskes et al., 1994). Quantitative measurements for model confirmation include RMSE, slope and intercept from the regression between simulated and observed values (Hunt and Boote, 1998), and mean error (ME),

$$ME = \frac{1}{n} \sum_n (x_i - y_i) \quad (5)$$

which evaluates model bias.

### **Molecular Marker Length Polymorphisms Linked to *E* Loci and Cultivar Genotyping**

Microsatellites or simple sequence repeats (SSR) are highly informative and polymorphic genetic markers in soybean (Akkaya et al., 1992). These markers are composed of tandemly repeated 2-5-nucleotide DNA core sequences of different forms such as (CA)<sub>n</sub>, or (GT)<sub>n</sub>, being (AT)<sub>n</sub>/(TA)<sub>n</sub> and (ATT)<sub>n</sub>/(TTA)<sub>n</sub>, the most frequently found in the soybean genome (Rowgen et al., 1995; Akkaya et al., 1992). Because sequences flanking the SSR are highly conserved within individuals of the same species, primers for polymerase chain reaction (PCR) can be designed. The amplification of the tandem repeats in different genotypes can yield product length differences due to differences in the number of repeats. SSRs can be used for identification of genotypes, or as markers for a given locus conferring a particular phenotype. Because SSRs are locus specific markers with multiple alleles, they are more appropriate for genotyping soybean varieties than RFLP markers, which have multiplicity of loci and can make the genotyping ambiguous (Cregan et al., 1999; Rongwen et al., 1995)

SSR markers have been incorporated into the integrated (classical and molecular maps) soybean linkage map (Akkaya et al., 1995; Cregan et al., 1999). Based on published linkage reports, Cregan et al. (1999) assigned all but one of the classical linkage groups to a corresponding one in the molecular map. This allowed us to select SSR marker loci for loci *E*1 and *E*3, which have been placed on the classical map. SSRs linked to loci *E*2 and *E*4 were selected from maps constructed by Cregan et al. (1999) and Jun Abe et al. (2003), respectively. Primers for the selected SSR markers were used to screen several NILs differing at the *E*1 to *E*4 loci. This procedure led to the identification

of polymorphisms at the *E*-linked marker loci. Polymorphic SSR were used for cultivar genotyping.

### **DNA Extraction**

Seven near isogenic lines were selected to investigate the presence of length polymorphism of SSR at the loci *E1*, *E2* and *E3* (Table 3-1). Plants were grown in a greenhouse for two weeks. Soybean DNA was isolated from upper node leaves by a modified procedure of Murray and Thompson (1980) as reported in Vallejos et al. (1992).

Seven soybean cultivars: Yale, Williams 82, Vinton 81, Savoy, Omaha, Nile and Linford, were genotyped at *E1* to *E4* loci. This soybean germ plasm is from GRIN Germplasm-Soybean Collection and provided by Dr. Randal Nelson. DNA was extracted from 50 mg seed tissue flour in 150  $\mu$ L TES (Tris 0.1 M, pH 8; EDTA 5 $\mu$ M; NaCl 50mM) and 800  $\mu$ L of 1.25X extraction buffer (Tris.HCl 125 mM pH 7.8; EDTA Na 12.5 mM, pH 8.0; NaCl 1.4 M; CTAB 1.25%; NaSO<sub>2</sub> 0.5%). Samples were incubated for 50' at 65°C, and then extracted with chloroform-octanol (400  $\mu$ L). Phases were separated by centrifugation at 13000 RPM for 15'. Due to high concentration of polysaccharides and oils, this step was repeated 2-3 times. DNA was precipitated in isopropanol (600  $\mu$ L) for 30', incubated for an hour in 76% ethanol-Na acetate (0.2M), rinsed in ethanol 76%-NH<sub>4</sub> acetate (10 mM) for 30'' and air dried for hour. The pellet was resuspended in 800  $\mu$ L of TE buffer (Tris.HCl 10 mM; EDTA.Na 1.0 mM).

### **Polymerase Chain Reaction (PCR) and PCR Product Separation**

Reaction mixes contained 1 X PCR buffer [20 mM Tris-HCl (pH 8.4), 50 mM KCl] (Cat. No. 10342-020, Invitrogen, CA), 1.5 mM MgCl<sub>2</sub>, 200  $\mu$ M of each nucleotide, 0.1  $\mu$ L of 300 Ci mmol<sup>-1</sup>  $\alpha$ -<sup>32</sup>PdATP, 0.1  $\mu$ M of 3' and 5' end primers, 0.5 unit *Taq* DNA polymerase and 30 ng of soybean genomic DNA in a total volume of 20  $\mu$ L. Primer

sequences for *Satt*227, *Satt*557, *Satt*581, *Sat*\_038, *Satt*229, *Satt*006, *Satt*513 and *Satt*496 are published elsewhere (<http://129.186.26.94/ssr.html>; Cregan et al, 1999).

Thermocycling consisted of a 30'' denaturation at 94°C, a 30'' annealing at 50°C, and a 30'' extension at 72°C for 35 cycles on a Geneamp PCR System 9600 (PerkinElmer, Inc). The thermocycle had an initial denaturation phase of 2' at 94°C, and a final extension phase of 5' at 72°C. Samples were denatured at 72°C in formamide for 5' and quenched on ice before loading. PCR products (3µL lane<sup>-1</sup>) were separated on a DNA sequencing vertical gel containing 6% Long Ranger (Cat. No. 50611, Cambrex Bio Science Rockland, Inc.), 0.5 X TBE (Tris 0.5 M, Boric acid 0.445 M, EDTA 10mM) and 6M UREA, at 50W constant power for 90 min. PCR amplification products were visualized by autoradiography on a Kodak X-OMAT film (Cat No. 1651512, Kodak).

#### **Predicting Time to Maturity of Public Cultivars of Soybean Grown in Variety Trials**

We used the gene-based model to simulate growth and development of a set of soybean public varieties grown in variety trial network in Illinois USA. The trial network consisted of eight locations: Belleville, Urbana, Dekalb, Dixon, Dwight, Monmouth and Perry, where soybeans were grown between 1995 and 1999. Yield and time to maturity data along with crop management data are available at (<http://vt.cropsci.uiuc.edu/soybean.html>). Weather data are from the Midwestern Regional Climate Center (<http://mcc.sws.uiuc.edu/>). Soil parameters were provided by Dr. T. Mavromatis (pers. comm.). Genetic coefficients controlling plant development were estimated as functions of *E* loci after genotyping, while remaining genetic coefficients were from Mavromatis (unpublished) using a calibration procedure described in Mavromatis et al. (2001).

## Results and Discussion

We incorporated gene action into the CROPGRO-Soybean model (Boote et al., 1998) using a two-step procedure; first we calibrated genetic coefficients in CROPGRO-Soybean, and second we developed linear models to predict these from *E* loci information, using data collected in a field experiment in 2001 (Chapter 2). The model was confirmed using a completely independent data set collected during the 2002 season, at the same location. For model development and confirmation we used a set of soybean near-isogenic lines (Table 3-1) that spans a wide range of life cycle and development phases durations (Fig. 3- 2). Life cycle varied between 67 and 132 days, the period of pod addition varied between 16 and 64 days, and pod number varied between 4 and 87 pods per plant. Statistical properties were similar for the data set used for model development (season 2001) and confirmation (season 2002); however, some differences were evident in extreme cases.

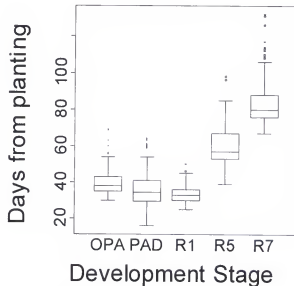


Figure 3-2. Genetic variability on soybean development within the set of near-isogenic lines grown in Gainesville during 2001 and 2002. These data were used to calibrate and confirm CROPGRO-Soybean. OPA: onset of pod addition, PAD: pod addition duration. R stages are defined in Fehr and Caviness (1977)

## Calibration and Evaluation of CROPGRO-Soybean

### Soybean development during 2001 season

CROPGRO-Soybean accurately simulated soybean phenology after calibration. RMSE in estimating plant development varied between 1.4 to 3.2 days (Table 3-3). The lower boundary in the RMSE is close to the observational error, defined by the interval between field observations. Figure 3-3 shows the relationships between observed and simulated values for time to flowering, time to maturity, time to onset of pod addition, time to last leaf on main stem node, and pod addition duration. There was generally good agreement between observed and simulated values, although this decreased as development progressed towards physiological maturity. This progressive decrease in predictive ability was observed before (Grimm et al., 1993; Grimm et al., 1994) and apparently is related to the propagation of errors as the simulation of later stages progresses and the uncertainty in measuring late stages of development.

Table 3-3. Comparative evaluation between CROPGRO-Soybean and CROPGRO-Soybean parameterized using *E* loci information using data collected during 2001.

Physiological Process	CROPGRO-Soybean				CROPGRO- <i>E</i> loci			
	RMSE	ME	$R^2$	$b^*$	RMSE	ME	$R^2$	$b$
Time to flowering	1.4	-0.17	0.92	0.89 (0.05)	2.7	-0.1	0.77	0.7 (0.08)
Time to onset of pod addition	2.5	0.51	0.88	1.0 (0.07)	4.0	0.32	0.72	0.83 (0.1)
Time to last mainstem node	3.2	0.69	0.93	1.05 (0.06)	4.2	1.14	0.87	0.99 (0.08)
Time to maturity	2.5	-0.45	0.96	1.08 (0.05)	3.9	-0.25	0.87	0.99 (0.08)
Pod addition duration	10.3	-9.6	0.46	1.04 (0.2)	9.8	-9.0	0.51	1.02 (0.19)

\* slope of the linear regression between simulated and observed values. Standard errors shown in parenthesis ( $df = 27$ ).

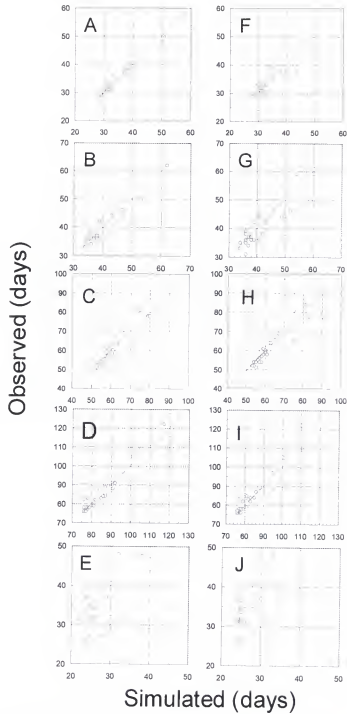


Figure 3-3. Relation between observed and predicted (A-F) time to flowering, (B-G) time to onset of pod addition, (C-H) time to last mainstem node (R5), (D-I) time to physiological maturity, and (E-J) pod addition duration. (A,B,C,D,E) Soybean development simulated using CROPGRO-Soybean and genetic coefficients calibrated for each NIL. (F,G,H,I,J) Soybean development simulated using CROPGRO-Soybean with genetic coefficients estimated from *E* loci using equations in Table 3-4. Data collected during 2001.

The calibration procedure provided genetic coefficient estimates with little or no bias as reflected in ME, and a slope not significantly different from one (Table 3-3). The model accounted for 88% to 96% of the variance in observed values (Table 3-3), which were within the range of results reported before for soybean (Mavromatis et al., 2001; Mavromatis et al., 2002; Grimm et al., 1994; Elizondo et al., 1994) and beans (Hoogenboom et al., 1997; White and Hoogenboom, 1996).

### **Simulation of pod addition duration and pod number in 2001**

The model systematically underestimated the duration of pod addition (Table 3-3). The pattern in the simulation error did not vary significantly between planting dates ( $P > 0.05$ ) suggesting the underestimation of pod addition duration is not related to the parameterization of the photoperiod sensitivity. Alternatively, the parameter PODUR could be underestimated leading to shorter pod addition duration. However, the value of PODUR is close to the maximum value estimated for standard cultivars in CROPGRO-Soybean (Jones et al., 2003). Furthermore, increasing the value of PODUR did not prevent the model to underestimate pod addition duration.

However, Colson et al. (1995a) showed that SOYGRO V5.42 (Jones et al., 1989) simulated adequately pod addition for different varieties of maturity groups from 00 to II. Two model parameters that most affected pod addition in Colson et al. (1995a) study were SDPDVR (average number of seeds per pod), and SDVAR (seed growth rate,  $\text{mg d}^{-1}$ ). The average number of seeds per pod, seed weight and seed fill duration were maintained constant due to the common genetic background of the near isogenic lines and the lack of genetic evidence linking quantitative trait loci for seed size and the *E* loci.

Despite the underestimation of pod addition duration, and in agreement with Colson et al. (1995a) results, CROPGRO-Soybean simulated well the number of pods per

plant. The slope of the regression between simulated and observed values was not significantly different from 1 ( $\alpha=0.05$ ) and the intercept was not different from zero. Because the model underestimated pod addition duration, one can reason that CROPGRO-Soybean overestimate the rate of pod addition. However, PODUR was set close to the maximum reported values (Jones et al., 2003), which correspond to a slow rate of pod addition. Alternatively, one can hypothesize that the model simulates well the rate of pod addition and pod addition duration early in the reproductive period, when the rate of pod addition is highest, but simulates poorly the slow pod addition late in the reproductive period. Data collected in the field was based on the last pod formed, even though late pods were added at a lower rate during the late reproductive period.

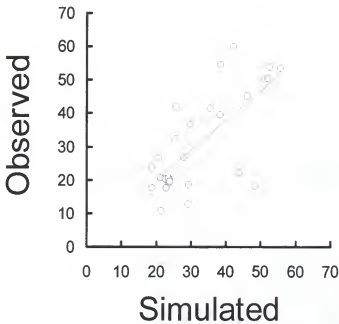


Figure 3-4. Observed and simulated pod number per plant for experiments conducted in Gainesville in two planting dates during 2001.  $y = 0.91 (\pm 0.17)x + 1.4 (\pm 5.9)$ ;  $R^2 = 0.52$

Colson et al. (1995a) showed that cultivars Weber, Argenta and 86-07 had a very slow change in pod number after R5. Furthermore, they observed a slow increase in pod

number until R7 for cultivar Argentina. Our measurements for end of pod addition were 15 days later than R5 on average indicating the existence of a period of slow pod addition. This observation supports the hypothesis that the model underestimates pod addition duration yet simulates well pod number.

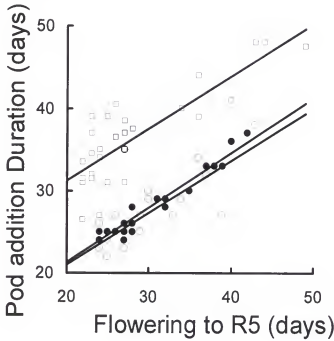


Figure 3-5. Relationship between R5 and the duration of pod addition. R5 stage is an estimator of the onset of sink development (first seed). (•) indicate values simulated using CROPGRO-Soybean [ $y = 0.63 (\pm 0.07)x + 8.4 (\pm 2.2)$ ;  $R^2 = 0.74$ ], (o) indicate values simulated with CROPGRO-Soybean and genetic coefficients estimated from *E* loci information [ $y = 0.67x (\pm 0.07) + 7.9 (\pm 2.3)$ ;  $R^2 = 0.74$ ], and (□) indicate observed values [ $y = 0.63 (\pm 0.06)x + 18.71 (\pm 2.2)$ ;  $R^2 = 0.72$ ]. Simulations conducted for the set of NILs listed in Table 3-1 grown in Gainesville in two planting dates during 2001 (Chapter 2).

Pod addition duration and time to R5 are linked in CROPGRO-Soybean through the simulation of carbon and nitrogen allocation to pods and seeds, and are highly correlated in field conditions (Chapter 2). Experimental results showed that pod addition duration and the duration from flowering to begin seed growth were both correlated with seed number (Kantolic and Slafer, 2001; Egli and Bruening, 2000) suggesting a co-

regulation between these processes. Figure 3-5 shows that CROPGRO-Soybean can reproduce this pattern for the set of near isogenic lines listed in Table 3-1 in two planting dates grown during 2001. The slopes of the regressions between the duration from R1 to R5 and pod addition duration calculated for simulated and observed values were not significantly different ( $\alpha=0.05$ ). There were significant differences ( $P<0.05$ ), however, between the offsets of these regressions due to the model underestimation of pod addition duration. These results further support the hypothesis that CROPGRO-Soybean simulates well the rate of pod addition and pod addition duration early in the reproductive period, as shown by the adequate simulation of pod number (Fig. 3-4) and the strong correlation between pod addition duration and time to R5 (Fig. 3-5), but simulate poorly the slow pod addition late in the reproductive period.

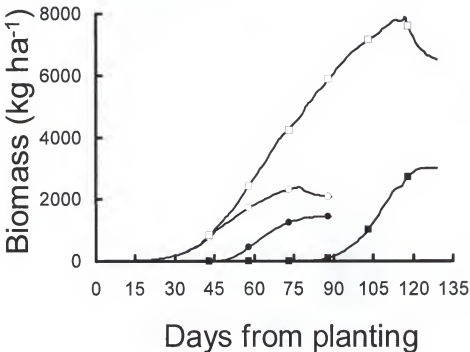


Figure 3-6. Simulated biomass accumulation for the near-isogenic lines L98-2064 ( $\square$ ) and L92-21 ( $\circ$ ) grown in Gainesville in 2001. Open symbols indicate total biomass and close symbols indicate mass in seed. Lines were simulated using CROPGRO-Soybean using genetic coefficients estimated using 2001 data.

### **Simulation of plant development effects on the dynamics of biomass accumulation, yield and harvest index**

Figure 3-6 illustrates the effects of variations in genetic coefficients controlling plant development on the dynamics of biomass accumulation, yield and harvest index of two near-isogenic lines of contrasting characteristics. For the longer season NIL, CROPGRO-Soybean simulated a longer reproductive period, a higher biomass production and yield in pods, but a lower harvest index than for the shorter season NIL. Previous studies have shown that short season cultivars have lower yield but higher harvest index (Kumudini et al., 2001; Spaeth et al., 1984).

### **Estimating Genetic Coefficients from Genotypes**

Genetic coefficients were estimated from *E* loci with different levels of accuracy (Table 3-4). The proportion of the total variance explained by the linear models varied between 32% and 88%, which are similar the values reported for the relationships in Genegro (White and Hooogenboom, 1996). All loci, coded as the number of dominant alleles for modeling purposes in a variable named *NLOCI*, affected the coefficients CSDL, PPSEN and SD-PM. Because CSDL and PPSEN mediate the direct influence of photoperiod on rate of development, their relationship with *E* loci was expected and was consistent with previous models of time to flowering based on the action of *E* loci (Stewart et al., 2003; Upadhyay et al., 1994a). However, the relationship between *E* loci and SD-PM was not that accurate and straightforward, indicating that photoperiodic effects are indirectly amplifying differences in phase duration. Therefore, this modeling exercise helped identify two modes of action of *E* loci on soybean development. One mode acts by the modulation of the critical photoperiod and photoperiod sensitivity. A

second mode regulates the number of physiological days required for a phase to complete.

Specific *E* loci were observed. *E1* alone showed major control on PPSEN, EM-FL, V1-JU and SD-PM, and in interaction with other loci, on PPSEN. This result confirmed the hypothesis that *E1* affected the juvenile phase as inferred from Upadhyay et al. (1994b). It also confirmed previous evidence of epistatic effects between *E1* and *E3* on the regulation of time to flowering (Upadhyay et al., 1994a). Notably, *E1* had a negative effect on physiological days from first seed to maturity, consistent with previous observations indicating that *E1* hampers soybean development during the reproductive period (McBlain et al., 1987).

Table 3-4. Associations between *E* loci and genetic coefficients in CROPGRO-Soybean. Dominant and recessive alleles take values of 1 and 0, respectively. *NLOCI* denotes for the number of dominant alleles<sup>†</sup>.

Genetic Coefficient	Linear Model	R <sup>2</sup>
CSDL	$CSDL = 14.33 - 0.44 NLOCI + 0.27 E3 - 0.48 E5 + 0.18 NLOCI E5$	0.88
PPSEN	$PPSEN = 0.11 + 0.063 NLOCI + 0.58 E1 - 0.13 E1 NLOCI$	0.70
EM-FL	$EM-FL = 20.77 + 2.1 E1 + 1.8 E3$	0.78
FL-SD	$FL-SD = 0.56 FL-VS$	--
FL-VS	$FL-VS = 20.9 + 0.67 NLOCI$	0.47
SD-PM	$SD-PM = 35.2 - 1.0 NLOCI - 9.2 E1 + 2.0 NLOCI E1$	0.57
V1-JU	$V1-JU = 4.16 E1$	0.71
R1PPO	$R1PPO = 0.1 + 0.066 NLOCI$	0.32

<sup>†</sup> FL-SH, FL-LF and PODUR did not significantly varied with *E* loci, they were set equal to 5.0, 26 and 13 photothermal days respectively for all NILs

Locus *E5* showed an association with CSDL (Table 3-4). In previous experiments it was shown that *E5* has major control over the duration of pod addition (Chapter 2). This simulation study shows that the effects of *E5* on pod addition duration by its effects on the delay on the transition of the apex from vegetative to reproductive. Longer pod addition duration would be the result of a longer time to sink development allowing the

plant to set more pods. A strong interaction was also shown between *E3* and *E5* on regulating pod number (Chapter 2). The results here suggest that the interaction between *E3* and *E5* is via their effects on CSDL. A reduction in CSDL would delay the onset and rate of seed growth under long photoperiods, increasing pod addition duration and seed number.

**Evaluation of CROPGRO-Soybean with Genetic Coefficients Estimated from *E* Loci. Data collected during 2001 season.**

CROPGRO-Soybean simulated well (RMSE = 2.7 to – 4.0 days) soybean phenology when genetic coefficients were estimated from near-isogenic lines genetic makeup (Fig.3- 3, Table 3-3). Figure 3-3 shows the relationships between simulated and observed values for different phases of soybean development. The RMSE increased and the proportion of the explained variance of the observations decreased relative to the calibrated results, which did not use *E* loci information. This result is expected due to the propagation of errors intrinsic to linear models used to estimate genetic coefficients, although this characteristic was not observed in the development of Genegro (White and Hoogenboom, 1996).

Predictions of time to onset of pod addition and pod addition duration using genetic coefficients estimated using *E* loci information showed lower bias than the predictions using CROPGRO-Soybean with coefficients calibrated to the original data. With the exception of time to flowering, all predictions showed good agreement with observed results with little deviations from the 1:1 line (Fig.3-3; Table 3-3). The model insensitivity in predicting time to flowering when parameters were estimated from genotypes ( $b < 1$ ;  $P < 0.05$ ) was due to one extreme value (50 days); which upon removal,

the slope of regression between observed and simulated values was not significantly different from one.

As shown in the previous section, CROPGRO-Soybean underestimated pod addition duration. This problem persisted when the genetic coefficients were estimated from *E* loci, and it has been discussed. The model when parameterized using *E* loci information can reproduce the relationship between time to R5 and pod addition duration. This result support the hypothesis that CROPGRO-Soybean may lack of adequate mechanisms to simulate the slow pod addition during the late reproductive period. As shown by Colson et al. (1995a) there is small contribution of these cohorts to total pod number.

#### **Model Evaluation with Independent Data from 2002**

We tested the model capabilities to predict crop development using an independent data set collected during 2002. The model accurately predicted reproductive development with low bias. RMSEP (root mean square error of prediction) ranged between 2.6 to 7.5 days, and MEP varied between 5.9 to -1.1 days (Table 3-5). These values are slightly higher than RMSE of the calibration (Table 3-3) data set. Despite the small decrease in the model precision relative to the calibration values, these RSMEP are comparable in magnitude with the precision of the measurements (2-4 days) and previous modeling results (Grimm et al., 1993).

CROPGRO-Soybean, run either with genetic coefficients estimated using data collected during 2001 or estimated from *E* genotypes, showed poor sensitivity when predicting the onset of pod addition and physiological maturity. The slope of the regression between observed and simulated values was different from one ( $P < 0.01$ ) (Table 3-5). Systematic deviations were observed in the predictions of physiological maturity

(Table 3-5). The model consistently under-predicted the time of physiological maturity for those long cycle NILs grown in early plantings, while errors for late plantings were not systematic.

Table 3-5. Confirmation of CROPGRO-Soybean and CROPGRO-*E* loci to predict independent observations (2002) of soybean development.

Physiological Process	CROPGRO-Soybean				CROPGRO- <i>E</i> loci			
	RMSEP <sup>‡</sup>	MEP	$R^2$	$b^*$	RMSEP	MEP	$R^2$	$b$
Time to flowering	2.6	0.1	0.65	0.79 (0.11)	2.8	0.38	0.62	0.73 (0.12)
Time to onset of pod addition	2.7	0.69	0.85	0.73 (0.06)	2.9	0.55	0.81	0.71 (0.07)
Time to last mainstem node	3.31	0.34	0.91	0.97 (0.06)	3.8	1.11	0.89	0.97 (0.07)
Time to maturity	7.6	-3.7	0.88	1.42 (0.09)	7.3	-5.9	0.89	1.39 (0.09)
Pod addition duration	13.0	-10.1	0.71	2.3 (0.28)	12.7	-10.4	0.68	2.2 (0.29)

\* slope of the linear regression between simulated and observed values. Standard errors shown in parenthesis ( $df = 27$ )

<sup>‡</sup>RMSEP denotes root mean square error of prediction since it was calculated using data collected during 2002, which were not used to estimate genetic coefficients, or to fit the relationships between *E* loci and genetic coefficients.

After removing six simulations for early plantings and long life cycle (>105 days) the model predicted physiological maturity without systematic bias ( $b=1$ ;  $\alpha=.01$ ). The removed data points had values greater than the upper quartile plus 1.5 times the inter-quartile range (Fig. 3-2). The simulation error can arise from measurement errors in R7.

### SSR Length Polymorphisms Linked to *E* Loci and Cultivar Genotyping

Polymorphisms were detected at several *E*-linked SSR marker loci in the NILs (Fig.3- 7). Results obtained with this survey support the proposed location of the *E* loci on the molecular map. Because of the linkage between the SSR and *E* loci position, we can infer with varying degrees of certainty the presence of the dominant allele in each public soybean cultivar. Due to the close linkage between *Satt557* and *E1*, the expected

uncertainty in determining the presence of the dominant allele is in the order of 1% since there is about 1.2 cM from *Satt557* to *E1* (Cregan et al., 1999; Jun Abe et al., 2003). The uncertainty increases for *E4*, which was mapped 5.0 cM apart from *Satt496* (Jun Abe et al., 2003).

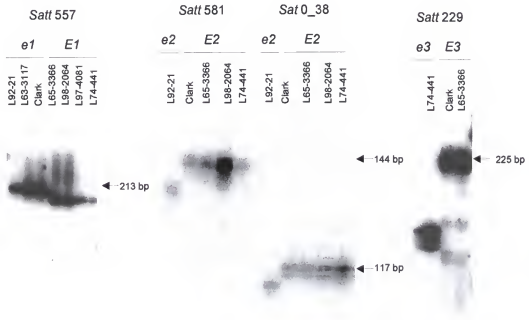


Figure 3-7. SSR length polymorphisms linked to *E* loci. Fragment size shown for Clark cultivar is from Soybase (<http://129.186.26.94/ssr.html>).

For a given SSR uncertainty, inferring the dominant *E2* is largest; the distances between *E2* and *Satt518* and *Sat\_038* were estimated as 17.2 and 18.3 cM, respectively. However, these markers are flanking the *E2* locus, in which case, the error of inferring the presence of *E2* when *E2* is absent due to double recombination is on the order of 3%. Uncertainty detecting *E3* can be as high as 12%. We were able to locate *E3* within a bracket of 14 cM between *Satt006* and *Satt513*. However, within this bracket only *Satt229* showed length polymorphism between the near isogenic lines.

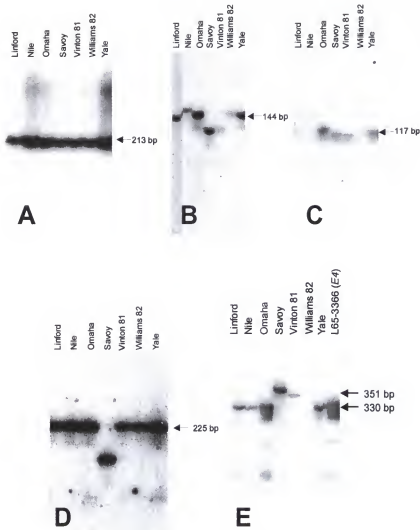


Figure 3-8. *E* loci genotypes for a set of soybean public cultivar. PCR fragment separation shown for *E1-Satt557* (A), *E2-Satt581* (B), *E2-Sat\_038* (C), *E3-Satt229* (D) and *E4-Satt496* (E).

The soybean cultivars varied at the *Satt* loci (Fig.3- 8). Given the uncertainty in the determination of each allele at a given locus, we inferred from the PCR fragment sizes the *E* loci makeup for each cultivar as follows, Linford *e1E2E3E4*, Nile *e1E2E3E4*, Omaha *e1E2E3E4*, Savoy *e1e2e3e4*, Vinton 81 *e1e2E3e4*, Williams 82 *e1E2E3E4*, Yale *e1E2E3E4*. At locus *Satt496* the PCR fragment sizes differ from previous reports (Jun Abe et al., 2003). *E4* alleles were determined by comparing the fragment size at the locus *Satt496* relative to the near-isogenic line for Clark carrying the dominant allele *E4*. We

assume that deviations from this size indicate the absence of *E4* must hence the cultivar have the *e4* genotype.

All genotypes had the recessive allele at the marker locus *Satt557*, therefore its genotype is *e1*, and probably *e7* since these loci are tightly linked (Cober and Voldeng, 2001). The cultivar Vinton81, however, has grey pod pubescence. The alternative allele, tawny color was found associated with earlier maturity (Cober and Voldeng, 2001). The locus *T* controlled this trait and is tightly linked to *E1* (1.4 cM) and *E7* (4.0 cM). Fragment size for *Satt557* suggests that the genotype is *e1*, hence *e7*. However, from the color of the pubescence we can infer the genotype as being *E7*.

### **Predicting Soybean Yield and Maturity in Variety Trials**

We calculated genetic coefficients using equations in Table 3-4 and genotypes estimated from SSRs (Fig.3-8). These coefficients were used in CROPGRO-Soybean to predict crop development and yield. Figure 3-9 shows that the model was able to predict general trends in maturity dates and yields varying from 1.5 to 5.0 Mg ha<sup>-1</sup>. The model predicted 75% and 54% of the observed maturity date and yield variances. These values are within the lower range of values obtained in previous modeling studies predicting yield and development in variety trials (Mavromatis et al, 2001; Mavromatis et al., 2002).

We can identify some causes contributing to the slightly higher prediction errors in our study. This study predicted genetic coefficients from *E* loci genotypes. Even when a model based on six loci could account for as much as 75% of the variance in maturity date, which demonstrates the importance of these loci on the regulation of soybean development, other loci are involved in the regulation of soybean development and yield and were not included in our model (Mansur et al., 1993; Mansur et al., 1996; Orf et al., 1999a,b; Tasma and Schoemaker, 2003; Tasma et al., 2001).

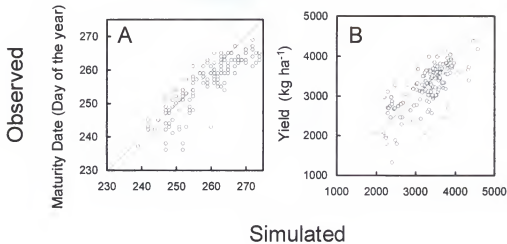


Figure 3-9. Simulated and observed time to maturity (day of the year) and yield ( $\text{kg ha}^{-1}$ ) of a seven soybean public varieties grown at eight locations and five years in Illinois (1995-99). Regression equations for time to maturity:  $y = -50.7 (\pm 9.7) + 0.8 \cdot x (\pm 0.04)$ ,  $R^2=0.75$ ; and yield:  $y = 689 (\pm 192) + 0.77 \cdot x (\pm 0.07)$ ,  $R^2=0.54$ . Time to maturity and yield RMSE were 5.2 (days) and 393 ( $\text{kg ha}^{-1}$ ) respectively. Yield RMSE is 12.3% of the average observed yields

The selection of terms in multiple regression linear models is an iterative process, which can lead to error in the definition of the model (Pineiro and Bates, 2000). Even though we used the nonparametric method CART (see Chapter 1) to gain confidence in the formulation of the linear models used to predict genetic coefficients from *E* loci, we cannot discard errors due to terms not included in the development of the regression equations. Errors can also be associated with the estimation of linear model parameters

Table 3-6. Prediction errors in time to maturity and yield for soybean public varieties grown in eight locations in Illinois (1995-99)

	Linford	Nile	Omaha	Savoy	Vinton81	Williams82	Yale
<i>Time to maturity</i>							
ME	2.7	2.6	3.2	-8.5	3.7	2.5	3.7
RMSE	5.0	4.6	5.2	10.0	6.0	5.1	5.5
<i>Yield</i>							
ME	129	108	-113	-390	373	173	-3.45
RMSE	355	359	329	537	567	485	338
Observed mean	3303	3110	3465	3611	2603	3160	3357

Bias and systematic errors of prediction of time to maturity suggests that these patterns can be due to errors in predicting certain genotypes, locations or years. As shown in Table 3-6, CROPGRO-Soybean significantly underestimated time to maturity for one variety of predicted shortest lifecycle but classified as maturity group II. This suggests that there is an error in the genotype based on *E* loci or that other loci regulate development in Savoy. It has been shown that other marker loci are involved in the regulation of soybean development and yield and were not included in our model (Mansur et al., 1993; Mansur et al., 1996; Orf et al., 1999a,b; Tasma and Schoemaker, 2003; Tasma et al., 2001).

We determined soybean genotypes based on the linkage between SSRs and the *E* loci. Recombination between loci can occur and there is a risk of inferring the presence of the dominant allele when it is absent. If this would have been the case for any of the marker loci, the genotype for Savoy would have been more sensitive to photoperiod, hence the model would have not underestimated the time to maturity and yield (Table 3-6). Error of prediction for Savoy illustrates well these limitations of the method. Recent advances in the identification and development of single nucleotide polymorphisms in soybean (Zhu et al., 2003) will help reduce the uncertainties associated with cultivar genotyping, and must hence increase the prediction skill of the model.

The statistics ME (eq.5) and RMSE (eq.4) calculated for time to maturity and all soybean cultivars but Savoy compares well with previous results (Table 3-3, Table 3-5, Mavromatis et al., 2001; Mavromatis et al., 2002). Errors in the simulation of time to maturity for cultivar Savoy propagated into the simulation of yield, which was underestimated as shown by the statistic ME equal to  $-390 \text{ kg ha}^{-1}$ . Simulated yield RMSE

for all but Savoy and Vinton81 cultivars varied from 9 to 15%, and ME varied between – 3 to 5% of average yields. I should be noted that yield predictions used a subset of genetic coefficients characterizing growth parameters not estimated from *E* loci but fitted by Mavromatis (pers. comm.). Model skill to predict yields and time to maturity can be considered acceptable for applications of crop models in agricultural production.

### Conclusions

A gene-based model for soybean was developed by incorporating gene action into the CROPGRO-Soybean model (Boote et al., 1998). This represents an advance with respect to previous models in predicting time to flowering from *E* loci information. The interaction of growth and development processes in CROPGRO-Soybean allows one to study the effects of genes controlling development on other physiological processes and traits of agronomic interest, not possible when modeling processes alone (Fig.3- 6). Despite the different mathematical approach used for modeling soybean development in this study relative to previous models for time to flowering (Cober et al., 2001; Upadhyay et al., 1994a; Stewart et al., 2003), CROPGRO-Soybean accurately predicted time to flowering and post-flowering development phases (Fig.3- 3; Table 3-3; Table 3-5). The prediction skill showed by CROPGRO-Soybean linked to *E* loci was comparable with that of Genegro for dry bean (White and Hoogenboom, 1996). However, systematic errors of post-flowering predictions in early plantings, probably associated with the parameterization of the temperature function  $f(T)$ , were identified (Table 3-5).

A genetic approach based on the use of near-isogenic lines was developed for model parameterization. This is a new approach for model calibration. This method allows the testing of processes and hypotheses underlying simulation models

independently. This study confirmed and strengthens our confidence in the approaches used to model development in CROPGRO-Soybean.

The model was able to reproduce not only final results of the interaction of physiological processes, such as pod number, but also showed skill in predicting the processes leading to the final outcome, such as the association between pod addition duration and the time to seed growth (Fig.3- 5). However, we identified that the CROPGRO-Soybean may lack adequate mechanisms to simulate late pod addition at very slow rates. We provide evidence that the model can adequately simulate pod number and the relationship between pod addition duration and time to R5, but underestimated pod addition duration, as measured in the field.

The gene-based model proved useful to understand an intrinsically complex biological system. Previous research indicated interactions between loci *E3* and *E5*, leading to the hypothesis of gene-gene interaction. The systems approach embedded in a crop model allowed the formulation of an alternative hypothesis of indirect gene-gene interaction through their independent effects on physiological processes. Most agronomic traits are quantitative in nature, polygenetically controlled and show strong GxE interactions. Gene-based models as shown in this study and others (Hoogenboom et al., 1997; Chapman et al., 2003) can help one understand and exploit gene-gene and gene-environment interactions.

Application of gene-based models relies heavily on their ability to adequately predict crop growth and development. Previous studies evaluated the models for their ability to reproduce the same data used in model development (Stewart et al., 2003; Cober et al., 2001; Upadhyay et al., 1994a; White and Hoogenboom, 1996). Others

simply made assumptions that gene action relationships would hold under the new environmental conditions (Chapman et al, 2003). For the first time a gene-based model was tested for its ability to reproduce yields and development at the field scale based only on the genetic makeup of the cultivar. Because prediction errors using a gene-based approach are comparable with those using conventional parameter estimation, gene-based models are a real alternative for yield simulation. Failure to simulate yield for Savoy and Vinton81 cultivars shows there is potential for improvement and thus to reduce uncertainties, errors and risks involved in the development and implementation of gene-based models.

Crop models, in contrast to other emerging tools in computational systems biology, integrate knowledge across disciplines and scales. This allows us to make inferences and study effects of genes at the organism level instead of the cell or molecular pathway (e.g., Davidson et al., 2002). The mechanistic approach used in CROPGRO-Soybean can considerably improve our understanding of the biological system relative to statistical approaches (e.g., Stoll et al., 2001). The successful simulation of soybean yield in variety trials supports this notion and encourages further research to integrate knowledge at molecular and organism levels. For the same reason and from a model application perspective, gene-based approaches can help reduce the requirements for expensive and intensive experimentation to provide up-to-date genetic coefficients. Gene based approaches can be significantly improved by the identification and incorporation of QTL regulating important physiological processes. Increasing the density of markers around relevant QTLs will reduce uncertainties during genotyping and can improve the simulation of crop traits.

## CHAPTER 4

### LINKING OPTIMIZATION ALGORITHMS AND GENE-BASED MODELS FOR CROP ENGINEERING IN TARGET ENVIRONMENTS

#### **Introduction**

Plant breeding faces immense challenges resulting from increased food demand, expansion of agriculture to marginal and diverse production areas, disease pressure, climate variability and reduced genetic variability. The interaction between genetics and environment raises questions about the ability of current crop cultivars to cope with these new environmental challenges. The development of adapted cultivars requires that changes occur simultaneously in structure, physiology, reproduction and development traits (Paterson et al., 1991). The narrow genetic basis observed in some breeding programs, as in the US soybean germoplasm (Gizlice et al, 1994; Keim et al., 1992), may limit the potential to breed for such cultivars (Manjarrez-Sandoval et al, 1997; Kisha et al., 1998). Efforts to broaden the genetic base may be futile if the introduced germoplasm does not increase the variability of desired traits.

Advances in plant molecular biology promises to irreversibly change plant breeding the way we know it today. Functional genomics will help us understand the molecular basis and genetic regulation of plant traits (Somerville and Somerville, 1999; Somerville and Dangl, 2000). Plant transformation protocols allow us to incorporate in the plant genome those genes that regulate the desired traits, whether they belong to the same or different organisms. Molecular makers can assist plant breeding by locating desirable genes and identifying combinations of loci that regulate quantitative traits

(QTL) (Paterson et al., 1991; Lee, 1995). Interactions between QTL and the environment makes marker-assisted selection a difficult task (Paterson et al., 1991). Because yield response can vary in the multitrait-environment space, a method that integrates knowledge across disciplines, including genetic interactions, pleiotropic effects, and a strong physiological framework, is required for

- designing crops for target environments
- identifying and prioritizing physiological traits and the underlying genes contributing to yield maximization
- assessing the effects of genetic base and selection strategies on yield gains

Crop models have been used to design ideotypes for target environments even when the environment was determined by a given management (Hammer et al., 1996; Boote and Tollenaar, 1994; Kropff et al., 1995). Sensitivity analysis has been widely used for ideotype analysis to study the effects of plant traits contributing to yield maximization in several crops (Boote and Tollenaar 1994; Boote et al. 2001; Boote et al., 2003; Aggarwal et al., 1997; White, 1998; Hunt, 1993; Kropff et al., 1995). Individual or combinations of model parameters within known genetic ranges were varied to study yield variations. These simulation studies suggest the need for varying multiple traits to attain significant, albeit modest, increases in yield. To design ideotypes based on multiple traits, Aggarwal (1997) used Monte Carlo simulations to generate 600 cultivars of rice. Hammer et al (1996) expanded the concept by linking a sunflower model with a simplex algorithm to optimize crop traits and its management for a given environment.

However, the former strategies for ideotype design are adequate only if that the yield response surface to physiological traits is smooth, the initial set of model parameters or traits can lead to the global optima, and there are no epistasis and

pleiotropic effects. These conditions are rarely met (Royce et al., 2001) or would require previous knowledge about adapted cultivars (e.g., Boote et al., 2003). Thus, sensitivity analyses and simplex optimization most likely lead to local maximum yield. In addition, the former strategies require a high level of expertise in using crop models. Crop modelers rather than plant breeders conducted most of these studies. These limitations can become particularly important when introducing a crop into a new environment, for assisting breeding of new crops, and whenever strong gene by environment by management interactions are present.

Gene-based approaches to simulate crop growth and development can account for epistasis and pleiotropic effects, enhancing model capabilities for ideotype design (Boote et al., 2001; Boote et al., 2003). The application of these models, however, requires that crop physiological mode of action of the trait is well understood and quantified, and the ecophysiological model is sufficiently detailed to simulate interactions between traits and the environment (Hammer et al., 1996; Aggarwal et al., 1997 ; Boote et al., 2001). Such models recently became available for common bean (White and Hoogenboom, 1996), barley (Yin et al., 2003), sorghum (Chapman et al., 2003) and soybean (Chapter 3) allowing their application to assist plant breeding and study the risks involved in ideotype design using traditional crop models. For a reduced number of genes, the application of gene-based models reduces to solving a combinatorial problem, as shown in common bean (Hoogenboom and White, 1999). However, with an increasing number of traits, loci and QTLs evaluated for yield maximization, robust optimization algorithms resistant to initial conditions and local maxima are required. These algorithms should handle both continuous and discrete or categorical variables.

The objectives of this chapter are:

- to develop an approach for “unsupervised” ideotype design for target environments by linking a gene-based crop model and a global optimization algorithm
- to evaluate the approach by its capability to identify traits contributing to yield maximization in target environments
- to study the effects of genetic base breadth and selection pressure on yield gains
- to study the risks of ignoring epistatic and pleiotropic effects for ideotype design

### **Materials and Methods**

#### **Linking Crop Models and Optimization Algorithms to Assist Plant Breeding**

One can think of three schemes for linking crop models and optimization algorithms to assist plant breeding. The conventional approach will use conventional crop models to provide the physiological framework and optimization algorithms that drive the crop model to yield maximization by querying the genetic coefficient space. Once the optimal combination of genetic coefficients are determined, one can search for quantitative trait loci (QTL) or Mendelian loci that are associated with those genetic coefficients. These solutions can become hypotheses to be tested in field experimentation.

This approach implies assuming independence among genetic coefficients, ignoring both epistatic interactions between QTLs and pleiotropic effects. Hence, unfeasible solutions can be found, overestimating potential genetic gains under current knowledge and available genetic materials. To relax the pleiotropic effect assumption, the simulation of a given trait could be conditioned to various genetic coefficients. Boote et al. (2001) successfully implemented this strategy to simulate the effects of the gene *dt* controlling growth habit by simultaneously modifying PODUR, time between flowering and the differentiation of the last node on the main stem, and FL-LF (see Table 4-1 for

definitions). This approach can provide useful insights to guide future research in biology that could ultimately lead to improved cultivars. Solutions arising from applying this procedure will produce ideotypes defining an upper bound of potential genetic gains for a given environment and crop management.

Gene-based models, on the contrary, do not make assumptions about pleiotropic and epistatic interactions through the values assigned to the genetic coefficients (Chapter 3; White and Hoogenboom, 1996). When using gene-based models, the optimization algorithm queries the loci-QTL space rather than the genetic coefficient space. There is a clear gain in the realism of ideotypes obtained by using gene-based models, but not without adding constraints to the search. So far, available gene-based models include effects of loci that exert major control on plant development. However, other loci have been shown to regulate plant development (Table 4-2). By not considering these other loci to estimate genetic coefficients, the ideotype search yields a lower bound on the potential genetic gains. In a hybrid approach that uses gene-based models, the optimization algorithm searches throughout both the loci space and genetic coefficient space. This procedure can improve ideotype design but increases the uncertainty associated with lack of knowledge about genetic controls of some traits.

### **Optimization with Adaptive Simulated Annealing**

Optimization methods can be classified into two groups: local search methods and global search methods (Hart et al., 1998; Royce et al., 2001). They differ with respect to the number of iterations required to converge to an optima, their sensitivity to initial conditions and their ability to handle discrete variables and criteria for determining that the algorithm has found a global optima. Local search methods such as Nelder—Mead simplex and Powell's conjugate directions, converge faster than global search

methods to an optimum but they are sensitive to initial conditions, plateaus, ridges, and discontinuities, all of which could lead to local optima.

Table 4-1. Genetic coefficients, ranges of variation, and examples for four maturity groups

Genetic Coeff.	Description	Range	Probe Cultivars Maturity Groups <sup>a</sup>				
			III	IV	V	VII	
<i>Genetic coefficients that can be replaced by functions of E loci</i>							
CSDL	Critical Short Day Length below which reproductive development progresses with no daylength effect (h)	11.78 - 14.6	13.4	13.1	12.8	12.3	
PPSEN	Slope of the relative response of development to photoperiod with time (h <sup>-1</sup> )	0.0 - 0.5	0.28	0.29	0.30	0.32	
EM-FL	Time between emergence and flower appearance (photothermal d)	15.5 – 28.5	19.0	19.4	19.8	20.8	
FL-SH	Time between first flower and first pod (photothermal d)	4 – 10	6.0	7.0	8.0	10.0	
FL-VS	Time from first flower to last leaf on main stem (photothermal d)	9-26	26	26	9	9	
FL-SD	Time between first flower and first seed (photothermal d)	10 – 17.6	14.0	15.0	15.5	16.0	
SD-PM	Time between first seed and physiological maturity (photothermal d)	26.0 - 38.7	34.0	34.5	35.0	36.0	
<i>Genetic coefficients not replaced by functions of E loci</i>							
FL-LF	Time between first flower and end of leaf expansion (photothermal d)	15 – 30	26.0	26.0	18.0	18.0	
SLAVR	Specific leaf area of cultivar (cm2/g)	175 – 400 <sup>†</sup>	375	375	375	375	
SIZLF	Maximum size of full leaf (cm2)	140 – 248 <sup>†</sup>	180.	180.	180.	180.	
WTPSD	Maximum weight per seed (g)	0.04-0.359 <sup>§</sup>	0.19	0.19	0.18	0.18	
SFDUR	Seed filling duration for pod cohort (photothermal d)	13.0 –56.0 <sup>  </sup>	23.0	23.0	23.0	23.0	
PODUR	Time required for cultivar to reach final pod load (photothermal d)	7 – 15	10.0	10.0	10.0	10.0	
V1-JU	Time required from first true leaf to end of juvenile phase (thermal d)	0 – 10 <sup>‡</sup>	0.0	0.0	0.0	0.0	
R1PPO	Increase in daylength sensitivity in post-flowering (h)	0.189 – 0.776	0.32	0.37	0.41	0.50	

<sup>‡</sup> From Boote et al. (2001); Boote and Tollenaar (1994); Mavromatis et al. (2001); Jones et al. (2003)

<sup>†</sup> Mian et al. (1998)

<sup>§</sup> Maughan et al. (1996), LeRoy et al. (1991), Mian et al. (1996)

<sup>\*</sup> Jones et al. (2003)

<sup>||</sup> Tomkins and Shipe (1996)

<sup>||</sup> Typically this value varies between 15 and 25 photothermal days

Table 4-2. Associations between marker loci, soybean traits and genetic coefficients.

Trait	Loci	Coefficient	References
First Flower	K365, K474_1, K474_2, A397-BLT029, A109_1, A385—G17_3, Satt6, A584-R079, Satt561, Satt365—Satt319, Sat_113, G173_1, Satt166, A489_1, Satt373, Satt150, Satt567, Sat_003, FT1, FT2, FT3, FT4, Satt205, Satt394, Satt380, Satt276, E1, E2, E3	EM-FL, V1-JUV	Keim et al., 1990; Mansur et al., 1993a; Mansur et al., 1996; Orf et al. 1999a,b; Yamanaka et al., 2001; Tasma et al., 2001; Chapter 2
First Seed	A397—BLT029, A385—G17_3	FL-SD	Mansur et al., 1993a; Chapter 3
Physiological Maturity	R183, BLT043, cr122, K472, A063, K365, K474_1, K474_2, A397, Satt79, R013_2, B032_2, R051, Satt6, O109, R079, A584, Satt432, Satt277, Satt561, Satt489, Satt150, Satt527, Satt373, Sat_003, Satt205, Satt380, E1, E2, E3, E4, E5, E7, Satt205, PHYB, FCA	EM-FL, FL-SH, FL-SD, SD-PM	Lark et al., 1994; Lee et al., 1996a,b, Keim et al., 1990; Mansur et al., 1993a,b; Mansur et al., 1996; Orf et al. 1999a,b; Tasma et al., 2001; Chapter 2; Chapter 3
Photoperiod Sensitivity	A397—BLT029, G8_15, A584—R079, L199_2, Satt489, Satt150, Satt567, K644_1, Drl, Satt373, Sat_077, G17_3	CSDL, PPSEN	Tasma et al., 2001, Jun Abe et al. 2003; Chapter 3; Tasma and Schoemaker, 2003
Seed Filling R1-R8	A109_1, R079, Satt561-Satt507, Satt508, T028_1, Satt291, Satt489, Satt150, Satt540, L050_14	SD-PM, R1PPO	Mansur et al., 1993a, Keim et al., 1990; Orf et al., 1999a; Mansur et al., 1996
Seed Number	T155, K443, A118, A059, A635, A262d, A257, BLT049_2, B031_1, A235_1, A816, K384, B166, Drl, Satt561, Satt187, Satt508, Satt174, Sat_036, Satt277, Satt150, Satt527, Sat_099, K001_1	---	Mansur et al., 1996; Orf et al. 1999a,b
Seed weight	A397—BLT029, BLT53NT3, A584—R079, A085, BLT049, K644_1	WTSD, THRESH, SFDUR	Mansur et al., 1993a; Maughan et al., 1996; Mian et al., 1996; Mansur et al., 1996; Orf et al. 1999a,b
Leaf Area	BLT043, A122D, A381, A089D_1, EV2E_1, A489, A169	SIZLF	Mansur et al., 1993a; Mansur et al., 1996; Mian et al., 1998
Specific leaf weight		SLAVAR	Mian et al., 1998

When local search algorithms are used, there is no assurance that the optimum is actually a global one. For difficult functions, one can try solving the problem several times from different starting points (Goffe et al., 1994; Jones et al., 2000). In contrast, global search algorithms, such as simulated annealing and genetic algorithms, are robust to local optima and discontinuities but require higher number of iterations (Corana et al., 1987; Goffe et al., 1994). Both types of algorithms have been used to solve optimization problems in agriculture (Thornton and McRobert, 1994; Hart et al., 1998; Mayer et al., 1998; Hammer et al., 1996), with global optimization algorithms becoming more frequently used due to the complex nature of the crop model solution space (Royce et al., 2001; Mayer et al., 1998; Hart et al., 1998).

Simulated annealing (SA) was proposed to solve large and complex functions in combinatorial optimization (Kirkpatrick et al., 1983). SA is based on random evaluations of the objective function (e.g., average yield) such that local optimum can be avoided. Corana et al. (1987) adapted the algorithm to optimize functions in a continuous domain; discontinuities in the function are allowed. However, Goffe et al. (1994) extended the Corana algorithm with checks for global optima and bounds to restrict the optimization to a subset of the parameter space. These characteristics of SA and their better performance relative to genetic algorithms (Goffe et al., 1994), makes SA an appropriate algorithm to use drive crop models for ideotype design.

SA starts by estimating the value for the objective function  $f$  at a given initial combination of parameters  $\mathbf{X}$ , an  $n$ -dimensional vector. A second evaluation  $f'$  is made at  $\mathbf{X}'$  by varying the  $i^{\text{th}}$  element,

$$x'_i = x_i + r \cdot v_i \quad (1)$$

where  $r$  is a uniformly distributed random number and  $v_i$  is the step length for parameter  $x_i$ . In maximization problems, if  $f'$  is greater than  $f$ , then  $X'$  is accepted replacing  $X$ , and the algorithm moves uphill. If this combination of parameters produced the largest value of  $f$ , then both  $X$  and  $f$  are recorded as the best current value of the optimum. When  $f'$  is lower or equal to  $f$ , the Metropolis criterion (eq.2) is used to decide acceptance of  $X$ . The Metropolis criterion is based on a simplified Boltzman probability distribution,

$$p = \exp( f' - f ) / \Phi \quad (2)$$

where probability  $p$  is compared with a uniformly distributed random number  $p_r$ . If  $p$  is greater than  $p_r$ , then  $X'$  is accepted and the algorithm temporarily moves downhill. Both the difference between function values and  $\Phi$  affects the probability of accepting downhill movements. At the beginning the user defines the parameter  $\Phi$  high enough such that there is a wide sampling of the function. As the optimization progresses,  $\Phi$  gradually decreases to,

$$\Phi' = r_\Phi \cdot \Phi \quad (3)$$

where  $r_\Phi$  [0,1] controls the rate at which the algorithm a) increases the probability of rejecting non-optimal steps, and b) narrows the search to the neighborhood of the current best solution. Low initial  $\Phi$  and  $r_\Phi$  can lead SA towards local optima. Adequate initial values for  $\Phi$  are such that the parameter space is fully sampled at the beginning of the simulation process. Values of  $r_\Phi$  greater than one will gradually increase  $\Phi$  and the breadth of the sample space. By setting  $r_\Phi$  to a value greater than one, inspection of parameter values for each  $\Phi$  helps identifying adequate initial values for  $\Phi$  for the given optimization problem. Corana et al. (1987) shows that a value of 0.85 for  $r_\Phi$  is adequate

to avoid local optima in complex problems. The algorithm ends by comparing the last  $N\varepsilon$  values for the largest function values, where  $\varepsilon$  denotes for a subjective small difference.

### **A Plant Breeding Metaphor for SA**

Some parallels can be established between plant selection in breeding and SA. Both processes are iterative, seeking to optimize an objective function, e.g. yield, by selecting best random combinations of traits, or parameters in SA. Traits, alleles and parameters are fixed during the breeding program and SA, respectively. The feasibility of finding global optimal solutions in SA is dependent on  $\Phi$ ,  $r_\phi$  and a set of boundaries. For example, the solutions of a search over the genetic coefficient space can be bound by the extremes shown in Table 4-1. Genetic gains in plant breeding depend on the breadth of the genetic base and the selection pressure. In SA terms, variable boundaries and  $\Phi$  are analogous to the breadth of the genetic base since they control the extent of sampling space, while  $\Phi$  and  $r_\phi$  are analogous to selection pressure by determining the probability of temporary acceptance of sub-optimal solutions (eq.2, eq3).

The analogies between SA and plant breeding suggest that a narrow genetic base can lead to local optima, hampering genetic gains and yield stagnation. It can be inferred from the parallel between processes that there is a tradeoff between the rate of genetic gain in the short term and yield stagnation due to local optimum in the long term. High selection pressure, or the rejection of a high fraction of sub-optimal solutions in SA, leads to rapid genetic gain in the short term. However, the pathway towards a global maximum yield may require the temporarily acceptance of suboptimal solutions. In the absence of this mechanism, both selection and optimization algorithms can lead to local optimum and yield stagnation in the long term. SA as a metaphor for a breeding process can be

useful to study the consequences of the breath of the genetic base and selection pressure on yield gains.

### Simulation of Crop Growth and Development

We simulated soybean yield and development using CROPGRO-Soybean (Boote et al; 1998). This dynamic process-oriented model incorporates the state of the knowledge of environmental and managerial effects on crop growth and development by simulating the effects of the environment on physiological processes, and on soil water and nutrient dynamics. Differences in morphological and physiological traits between genotypes are taken into account by a set of parameters named genetic coefficients (Hunt and Boote, 1998). A description of genetic coefficients, ranges and values for cultivars typically grown in the Pampas (herein named probe genotypes) are listed Table 4-1. The case study is described in more detail below.

Table 4-3. Soil parameters and soybean management for four locations in the Pampas

Location	Soil Properties <sup>†</sup>					Management <sup>‡</sup>		
	PESW (mm)	Depth (m)	CN	ISW (mm)	ISN (Kg ha <sup>-1</sup> )	Planting Date	Plant Density (m <sup>-2</sup> )	Maturity Group
Pilar	316	210	88	159	206	Nov 16 <sup>th</sup>	25	VI-VII
Pergamino	305	220	83	138	90	Nov 1 <sup>st</sup>	25	V
Santa Rosa	251	200	85	110	103	Nov 1 <sup>st</sup>	25	IV
Balcarce	206	120	80	40	29	Nov 15 <sup>th</sup>	35	III

<sup>†</sup> PESW: Plant extractable soil water, CN: runoff curve number, ISW: initial soil water, ISN: initial soil nitrogen. INTA researchers Dardanelli, Meira, Magrin and Travasso provided soil parameters for Pilar, Pergamino, Santa Rosa and Balcarce respectively.

<sup>‡</sup> Management practices obtained from AACREA, 1997

Differences between soils are characterized by variations in a set of parameters controlling, for example, runoff, soil water holding capacity and root growth. Ritchie (1998) provided a comprehensive description of soil water balance routines. Soil parameters, soil water and nitrogen content at the beginning of the simulation and typical management by location used for simulation are shown in Table 4-3. Daily weather data

are from the National Meteorological Service (Servicio Meteorológico Nacional) of Argentina and underwent intensive quality checking (Podesta, pers. comm.)

### Linkage between SA and CROPGRO-Soybean

The main SA program from Goffe et al., (1994) was modified to calculate the objective function value from simulated yield with CROPGRO-Soybean. A subroutine in FORTRAN was included to write the input files for CROPGRO-Soybean (SBGRO980.CUL and SBGRO980.ECO). This subroutine writes the files using either the values generated by the SA algorithm, constrained within a specified genetic range (Table 4-1) or from *E* loci combinations. In the latter case, genetic coefficients are calculated using corresponding equations (Table 3-4). Simulated yields are read from the SUMMARY.OUT file and the average for a given number of years is calculated and passed to the SA algorithm. Figure 4-1 shows the organization of the program and the flow of information within the code.

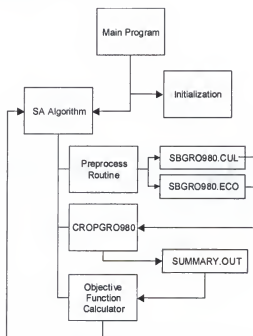


Figure 4-1. Representation of the linkage between SA algorithm and CROPGRO-Soybean.

The genetic coefficients FL-SD and FL-SH (Table 4-1) are specified independently in CROPGRO-Soybean input files. However, there must exist a minimum time between the onset of pod and seed growth. The minimum time between these two events is approximately six days. The “Big M” method (Ahuja et al., 1993) was implemented to avoid physiologically non-feasible solutions. When  $FL-SD \leq FL-SH + 6$ , simulated yield values were multiplied by 0.0000001 to reduce their probability of acceptance.

Table 4-1 lists the genetic coefficients that were optimized for ideotype design in all the case studies described below. The range for each parameter was defined based on previous research in crop modeling (Boote et al., 2001; Boote and Tollenaar, 1994; Mavromatis et al., 2001; Jones et al., 2003) and genetics (Mian et al., 1998; Maughan et al., 1996; LeRoy et al., 1991; Mian et al., 1996; Tomkins and Shipe 1996). For studying the effects of pleiotropic and epistatic effects on ideotype design (see case study below), the genetic coefficients CSDL, PPSEN, EM-FL, FL-SH, FL-VS, FL-SD, SD-PM, V1-JU and R1PPO (Table 4-1) were estimated using functions of E loci (Table 3-4).

### **Identification of Traits Contributing to Yield Maximization in Target Environments**

A wide range of soybean maturity groups varying from III to VI is grown in the Pampas. Most soybean production occurs between 31 S and 38 S and east of 64 W within a longitudinal annual rainfall gradient varying between 500 mm in the west to 1000 mm in the east (Hall et al., 1992). Similarly, there is an east-west gradient in soils varying from Entic Haplustols to Typic Argiudols, and soil water holding capacity (Table 4-3). The core production region is concentrated around the eastern location of Pergamino (Fig. 4-2). Recently, soybeans were introduced in cooler environments with Mediterranean rainfall regime around Balcarce (Fig. 4-2) and in the semiarid area near Santa Rosa.

We selected five environments in the Pampas to test the approach for ideotype design (Fig. 4-2a). The selected locations create a gradient of water stress for a small latitude range, and similar water stress conditions for a large latitudinal range (Fig. 4-2b). Because of decadal variations in rainfall in Santa Rosa, we expanded the range of water stress by selecting two periods of ten years each when precipitation was highest (1985-95) and lowest (1945-55). This set of target environments should suffice to identify plant traits for broad and specific adaptation.

For each of these target environments, we ran the program linking SA and CROPGRO-Soybean. In each SA iteration, CROPGRO-Soybean was run for ten years of daily weather; between 1985-95 for Balcarce, Pilar, Pergamino and Santa Rosa, and between 1945-55 for Santa Rosa. Table 4-3 describes the crop management for each location. The values were selected to represent typical practices in the region (AACREA, 1997). The SA parameter were set to  $\Phi = 25$  and  $r_\phi = 0.85$ . In each iteration, genetic coefficients were selected within known ranges of genetic variability (Table 4-1).

Yield maximization using SA was repeated three or four times using different combinations of initial conditions and random number generator seed numbers. Each realization consisted of an SA search for maximum yield following a different path. When all pathways converge to the same maximum, this can be considered a global maximum (Goffe et al., 1994). Genetic coefficients corresponding to this maximum yield defines the ideotype for a given target environment. We compared growth and development of each ideotype relative to the "cultivar" defined at the beginning of the optimization, and relative to a "probe" cultivar representative of the maturity group typically grown in each of the five environments (Table 2-1). These two comparisons

allowed us to evaluate the method for ideotype design and to assess the potential genetic gains for each location and a given management and set of traits.

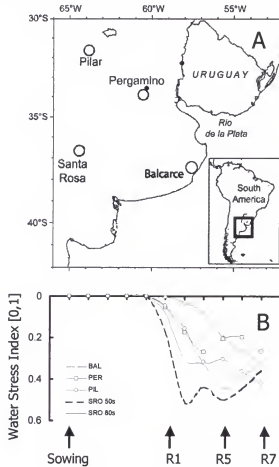


Figure 4-2. Case study location in the Argentine Pampas (A) and water stress index dynamics during the growing season (B). Water stress index and phenology was calculated using CROPGRO-Soybean for probe genotypes (Table 4-1). BAL: Balcarce, PER: Pergamino, PIL: Pilar, SRO: Santa Rosa.

### Genetic Base Breadth and Selection Pressure Effects on Yield Gains

We studied the effects of the breadth of the genetic base and selection pressure on genetic gains and yield stagnation due to local optima. Simulations were conducted for Santa Rosa during a wet decade (1985-1995), where the environmental challenge and potential genetic gains are largest. SA linked to CROPGRO-Soybean was run for six

combinations of  $\Phi$  (1, 5, 25) and  $r_\phi$  (0.01, 0.25, 0.85). When  $\Phi=1$  and  $r_\phi = 0.01$ , SA behaves similarly to a simplex algorithm; during the simulation only parameters that increase yields are accepted. In addition, the parameter space is explored in the neighborhood of the initial values set for each parameter. For  $\Phi=25$  and  $r_\phi = 0.85$ , approximately 50% of suboptimal solutions are accepted at the beginning of the simulation allowing a full exploration of the multidimensional parameter space.

The first step in this analysis is to demonstrate the existence of multiple local maximum. This is a necessary condition to test the hypothesis that high selection pressure and narrow genetic base can lead to yield stagnation due to local optima. Second, it must be demonstrated that the higher selection pressure or narrow genetic base leads to local maximum yield causing stagnation. If this hypothesis is false, we must observe a lack of a positive association between  $\Phi$  and  $r_\phi$  with simulated maximum yields.

### **Risks of Ignoring Epistatic and Pleiotropic Effects for Ideotype Design**

We studied the risks of ignoring epistatic and pleiotropic effects on yield maximization. This was a common assumption in previous research using crop models for ideotype design, in which genetic coefficients regulating growth and development were assumed independent (e.g., Paruelo and Sala, 1993; Aggarwal et al., 1997) We compared the results obtained in previous sections for Balcarce and Pergamino with new results obtained by driving the SA search throughout the  $E$  loci space instead of selecting values independently for each genetic coefficient. Equations derived in Chapter 3 Table 3-4) show that some  $E$  loci have pleiotropic effects, since they regulate different phenological phases, and epistatic effects, since they interact with each other to regulate photoperiod sensitivity. To understand the constraints imposed by pleiotropic and

epistatic effects on yield maximization we compared the genetic coefficients between optimal solutions and analyzed simulated growth and development.

## Results

### Convergence Towards a Global Maximum

SA solutions converged systematically to a global maximum. Figure 4-3 shows four realizations of the search process throughout the genetic coefficient space for Balcarce. Solutions converged to the global maximum of 3000 kg ha<sup>-1</sup> despite the pathway followed by SA. Differences between pathways are more evident early in the optimization when the yield increase was largest. The identification of the same maximum following alternative pathways confirmed that this value is a global maximum (Goffe et al., 1994). Similar results were obtained for other target environments whether SA search was performed over the genetic coefficient space or combined with the search over the *E* loci space.

Variability in simulated yields decreased with the number of runs, as expected, from a reduction in the value of  $\Phi$  and the increasingly narrower sample space from which genetic coefficients are withdrawn. However, the random variability close to the end of the search process was higher than expected relative to results obtained in other applications of SA (Kirkpatrick et al., 1983; Corana et al., 1987; Goffe et al., 1994; Ferreyra et al., 2002). The relative odd behavior of SA in this application arises from the implementation of the algorithm rather than the lack of convergence to a global maximum. When a random selection of a parameter is out-of-bounds, the algorithm selects at random a value within the genetic range rather than from the neighborhood of the best solution at the moment. Although this mechanism can increase the number of

necessary simulations, it can help prevent stagnation in a local optimum. Therefore, no attempt was made to modify the SA algorithm.

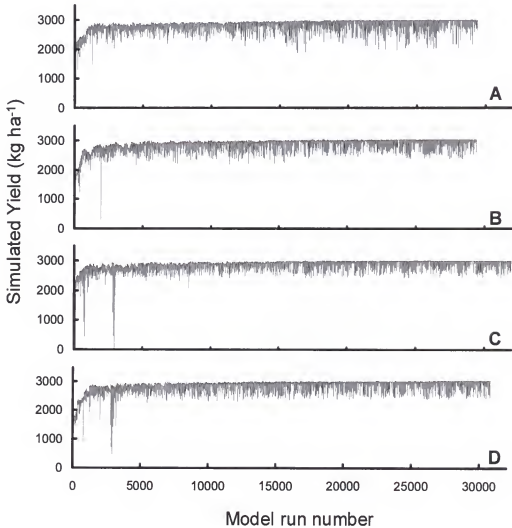


Figure 4-3. Evolution of simulated soybean yields in Balcarce during SA optimization. A-D are replications of the process from different starting conditions and seeds for the random number generator.

#### Physiological Analysis of Parameter Dynamics

The two-phase increase in yield is associated with four phases of modifications of the genetic coefficients. Within the first 5000 simulations with CROPGRO-Soybean, SA was able to identify solutions close the global maximum (Fig. 4-3). Regardless the initial combination of genetic coefficients, the largest increases in simulated yield occurred

within the first 2500 runs, after which yield increases were steady but at a low rate (Fig. 4-3).

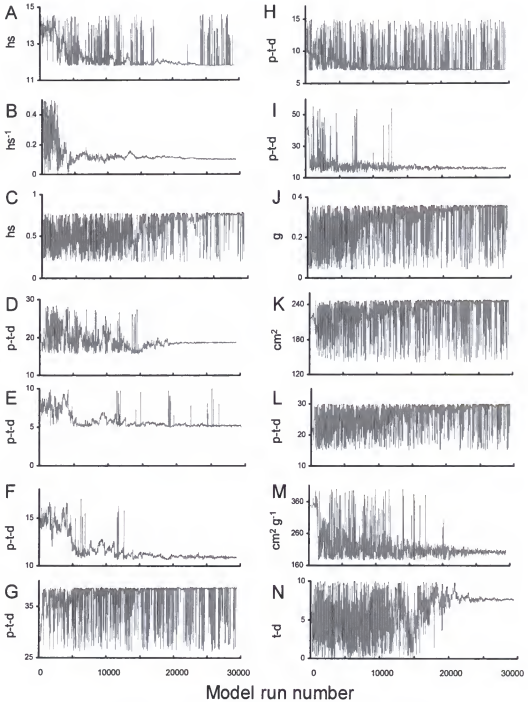


Figure 4-4. Genetic coefficient dynamics during SA optimization for Balcarce. See table 4-1 for definitions. A: CSDL, B: PPSEN, C: R1PPO, D: EM-FL, E: FI-SH, F: FL-SD, G: SD-PM, H: PODUR, I: SFDUR, J: WTPSD, K: SZLF, L: FL-LF, M: SLAVR, N: V1-JU

In a first phase, between runs 0 and 1000, yield increases can be explained by the rapid increase in seed filling duration (SD-PM) and seed growth rate. The latter is mediated by a reduction in the seed fill duration per seed (WPSD / SFDUR). There was also an increase in the synchronism of pod set through a reduction in PODUR (Fig. 4-4). These modifications led to a rapid increase in seed number per unit area increasing sink size and demand for photoassimilates.

In second phase ending around run number 2000, there is a relaxation of the source limitation induced in phase I. There is a sharp decrease in specific leaf area, an increase in duration of canopy expansion (FL-LF) and the duration between crop emergence and time to flowering (EM-FL). All these changes have the effect of increasing the vegetative mass, leaf area and canopy photosynthesis.

During phases I and II, the simulated ideotype is relatively insensitive to photoperiod due to a high CSDL. The optimization of plant development followed variations in photothermal duration of phenological stages. In contrast, in phase III, there was an increase in photoperiod sensitivity, particularly during the late post-flowering period when photoperiods are shortest, through a reduction in CSDL. However, to prevent the excessive duration of a) the growth cycle, which may increase risk of freeze damage, and b) the vegetative and early reproductive phases, which would reduce the duration of the reproductive period, there was a correlated reduction in PPSEN, EM-FL, FL-SH and FL-SD. The new intermediate ideotype had even longer seed fill duration, a reduced seed growth rate but increased leaf area duration that can support seed growth. The changes in the genetic coefficients caused a drastic change in yield component reducing the number of seeds per square meter and increasing weight per seed.

During the later phase, beginning around the run number 15000, photoperiod sensitivity increased during the reproductive period by increasing R1PPO. The consequent lengthening of the reproductive period was compensated by a shortening of the time to flowering. This was attained by increasing duration of the juvenile phase, which reduces the window during the EM-FL phase during which the plant is sensitive to photoperiod. The final ideotype had heavier seeds, which increased leaf area per plant during very early vegetative stages, and larger leaves, which increased early light interception. This increase in early vigor increased the source size (leaf area) that compensated for a reduction in the duration of the vegetative phase. Finally, higher sensitivity to photoperiod decreased seed growth rate decreasing the demand for nitrogen and its remobilization, hence, there was an increase in leaf area duration, seed weight and yield.

#### **Identification of Traits Contributing to Yield Maximization in Target Environments Ideotypes across a latitudinal range**

The SA optimization CROPGRO-Soybean identified ideotypes that yield more than both initial and probe genotypes across a latitudinal gradient (Table 4-4). Simulated yields for probe genotypes were in agreement with average yields recorded at AACREA farmers fields. For example, average soybean yield in CREA-Tandil was 2050 kg ha<sup>-1</sup> for the period 1988-96 (AACREA, 1997). That compares well with a simulated value of 2099 kg ha<sup>-1</sup>. Ideotypes outyielded probe genotypes by at least 40%, suggesting there is an important gap between current yield and yield in potential cultivars. However, as shown through this numerical experiment, genetic improvement would be realized by changing several traits, and some traits must be modified simultaneously.

Traits conferring the crop broad adaptation across locations were seed-fill duration, pod addition duration, time to flowering and photosynthesis. All these traits increased the partitioning of assimilates to reproductive structures increasing yield. Ideotypes for the three locations had longer seed fill and pod addition duration at an expense of a reduced time to flowering (Table 4-4). Increased photoperiod sensitivity during post-flowering relative to probe cultivars (Table 4-1, Table 4-5) caused a longer duration of the reproductive phase. This was accomplished by reducing CSDL, and by increasing R1PPO and photothermal requirements to complete seed fill duration (SD-PM). Fewer photothermal days to flowering in addition to longer juvenile phase relative to probe genotypes (Table 4-1; Table 4-5) decreased the sensitivity to photoperiod during the vegetative phase. Decreased specific leaf area increased leaf photosynthesis, and leaf area and biomass production (data not shown).

Table 4-4. Simulated soybean growth and development for a set of genotypes across a latitudinal gradient

	Balcarce			Pergamino			Pilar		
	INI <sup>†</sup>	OPT	Probe	INI	OPT	Probe	INI	OPT	Probe
Plant Development (days) <sup>t</sup>									
PAD <sup>‡</sup>	45	39	29	37	37	33.4	34	38	28.4
E-R1	118	44	49.4	44	38	50	37	24	59.7
R1-R5	41	23	28.4	34	81	37.8	30	26	40.3
R5-R7	34	56	37.4	31	44	41.4	30	52	39.4
Yield (kg ha <sup>-1</sup> ) and weight per seed (mg)									
Yield	131	3029	2099	1841	4859	2860	1709	3376	1581
SW	70.7	422	130	72	39.1	155	69	360	134

<sup>‡</sup>PAD: Pod addition duration; E: Emergence; LF: end of canopy expansion; R1 through R7 are soybean developmental stages according to Fehr and Caviness (1977). SW denotes weight of individual seed (mg)

<sup>†</sup> INI: genotype at the beginning of the optimization; OPT: genotype that maximizes yield; Probe: genotype of maturity group recommended for the location.

Specific adaptation strategies minimized negative effects of water stress during critical periods. The ideotype for Pergamino had high PPSEN, FL-SD and FL-SH relative

to the probe cultivar (Table 4-5, Table 4-1). These genetic coefficients determined a delayed onset of pod addition and seed fill such that these critical stages occurred during a period of lower water stress (Fig. 4-2b). In contrast, ideotypes for Pilar and Balcarce accelerated the onset of pod addition and seed fill to avoid or minimize the effects of terminal drought (Fig. 4-2b). In addition, ideotypes for Pilar and Balcarce set fewer seeds than probe genotypes, but seeds had higher weight, which further minimized the effects of terminal drought on yield. It is well known that water stress affects more severely seed number than seed growth rate, hence seed weight (Frederick et al., 1991).

Table 4-5 Initial and optimized genetic coefficients, yield and yield components for Pilar, Pergamino and Balcarce.

Genetic Coefficient	Initial Genotype			Optimized Genotype		
	Balcarce	Pergamino	Pilar	Balcarce	Pergamino	Pilar
CSDL	12.12	13.16	13.16	11.79	12.89	11.78
PPSEN	0.4	0.3	0.3	0.10	0.49	0.11
EM-FL	20	21	21	18.5	15.5	15.6
FL-SD	16	17	17	10.9	17.5	14.6
SD-PM	28	27	27	38.7	38.7	38.7
FL-LF	17.8	25	25	29.9	29.8	29.8
PODUR	14	12	12	7.0	7.02	7.0
SFDUR	40	40	40	15.9	18.8	16.6
FL-SH	6	6.4	6.4	5.22	9.44	8.95
SLAVR	350	350	350	200	175	175
WTPSD	0.25	0.22	0.22	0.358	0.045	0.354
SIZLF	220	199	199	246	241	173
JUV	1.1	3.0	3.0	7.7	9.9	8.9
R1PPO	0.2	0.3	0.3	0.74	0.77	0.61

Crop cycles for Pergamino and Pilar differed from the probe cultivars. Although we show potential increases in yield by modifying plant traits, a note of caution is in order. CROPGRO-Soybean does not account for losses due to either pests and diseases or harvest losses. Longer crop cycles as determined in this study can increase harvest losses in Pergamino due to water excess in late fall. Due to the monsoonal rainfall regime in Pilar, a shorter crop cycle as predicted can have associated higher yield losses due to

diseases such as phomopsis. However, there is some evidence that despite the fact that the most frequently planted maturity group in Pilar is group 7, maturity group 3 can yield more (Dardanelli, unpublished). Ultimately, the question is whether to avoid disease occurrence using long cycle cultivars or breed for cultivars of shorter cycle and disease resistance.

### **Ideotypes across a water stress gradient**

Yield for probe genotypes varied from 663 kg ha<sup>-1</sup> in the fifties in Santa Rosa to 2860 kg ha<sup>-1</sup> in Pergamino during the eighties, reflecting the magnitude of the yield limitations due to water stress. CROPGRO-Soybean driven by SA identified traits conferring broad and specific adaptation. As in the previous section, the greater fraction of the crop cycle dedicated to developing and filling reproductive structures increased yield in all target environments (Table 4-6). In Santa Rosa environments, however, this strategy was implemented mainly through variations in photothermal requirements rather than by changes in CSDL (Table 4-7). This parameter was set at 14 hs for the fifties environment and 13.5 hs for the environment in the eighties, showing little variation relative to the probe genotype (Table 4-1). Provided that ideotypes had a prolonged juvenile phase of 9.9 photothermal units, a value for CSDL around 14 h is high enough to confer reduced photoperiod sensitivity during the pre-flowering period. Increasing the parameter R1PPO from 0.37 to 0.76 h increased photoperiod sensitivity after flowering, thereby lengthening the duration of the reproductive period. As in previous cases, a reduced specific leaf area increased leaf photosynthesis.

There were significant variations in traits conferring specific adaptation between ideotypes designed for “dry” (1945-55) and “wet” (1985-95) environments (Table 4-7). The former had shorter duration to the onset of seed growth and end of canopy expansion

and smaller leaves. These genetic coefficients had a major impact on the development of leaf area.

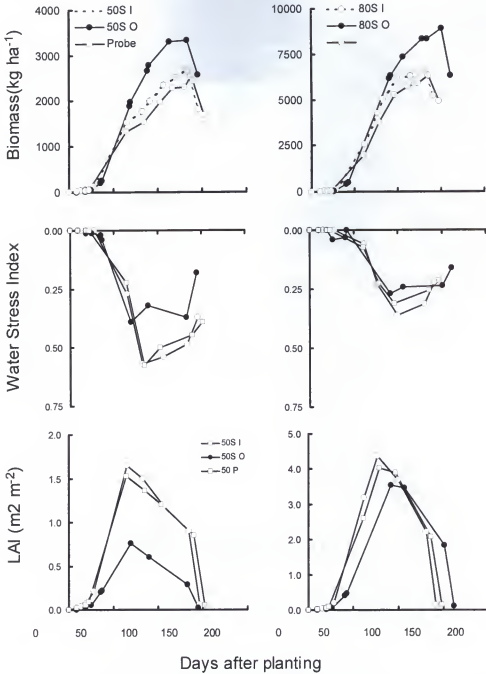


Figure 4-5. Simulated biomass, water stress index and leaf area index (LAI) dynamics for probe (P), initial (I) and optimized (O) soybean genotypes for Santa Rosa in contrasting environments of water availability. Left panels correspond to a dry scenario (1945-55). Right panels correspond to a wet scenario (1985-95). See Fig. 4-2 for differences in water stress index dynamics between scenarios.

Figure 4-5 compares biomass accumulation, leaf area index evolution and water stress index for the genotype at the beginning of the optimization, for a probe genotype and the ideotype. The ideotype had a higher rate and total production of biomass than the probe genotype. The higher biomass production and yield in the ideotype was attained despite the lower leaf area index. Although light interception was lower by the ideotype than the probe cultivar, the reduced transpiration maintained better soil water status, which offset the effects of lower light interception.

Table 4-6. Simulated soybean phenology for a set of genotypes under contrasting water stress environments in Santa Rosa. "Dry" 1945-55. "Wet" 1985-95.

	1945-55			1985-95		
	INI	OPT	Probe	INI	OPT	Probe
Plant Development						
PAD (days)	37	43	32	38	36	33
E-R1 (days)	56	28	55	51	32	51
R1-R5 (days)	41	52	38	38	64	35
R5-R7 (days)	27	43	36	31	44	39
Yield and Yield Components						
Yield (kg ha <sup>-1</sup> )	519	1206	663	1685	3240	2068
SW (mg)	82	40.6	134	90	33	153

Under a more favorable environment (1985-95), the ideotype also had a lower LAI relative to the probe genotype, which reduced water stress. This improved soil water use strategy allowed the extension of the growth cycle, and an increase in soil water availability during seed filling, thereby increasing solar radiation interception and biomass production (Fig. 4-5).

### Genetic Base Breadth and Selection Pressure Effects on Yield Gains

Genetic base breadth and selection pressure are embedded in the SA parameters  $\Phi$  and  $r_\phi$ . For a given level of  $r_\phi$ , maximum yield increased with increasing  $\Phi$  (Table 4-8). When  $r_\phi=0.01$ , there is drastic reduction in the probability of accepting suboptimal

solution after each update in  $\Phi$ . Therefore, when  $r_\phi = 0.01$  most of the variation in maximum yields was associated with the effects of  $\Phi$  on the breadth of the neighborhood around the parameters set as initial conditions explored by SA early in the optimization. This is analogous to the effects of genetic base breadth on yields. Adequate selection of  $\Phi$ , and by analogy, the selection of a wide genetic base is necessary for the identification of maximum yields close to the global maximum.

Table 4-7. Initial and optimized genetic coefficients, yield and yield components for Santa Rosa under two contrasting environments: 1945-55 and 1985-95

	EM-FL	FL-SH	FL-SD	SD-PM	SFDUR	WTPSD	FL-LF	SLAVR	SIZLF
Initial Genotype									
	21.0	6.4	17.0	27.0	40.0	0.22	25.0	350	199
Optimized Genotype									
1945-55	15.5	9.5	17.4	38.7	13.1	0.05	15.0	175	140
1985-95	15.5	9.6	15.6	38.7	21.7	0.05	28.9	175	238

We show evidence of existence of multiple local yield maxima in the parameter space determined by the selected genetic coefficients (Table 4-1) and under the physiological framework provided by CROPGRO-Soybean (Table 4-8). Simulated yields varied from 1685 kg ha<sup>-1</sup>, at the beginning of maximization with SA, to the global maximum of 3240 kg ha<sup>-1</sup> in Santa Rosa. Then, yield stagnation due to local optima can arise from a narrow genetic base breadth, a high selection pressure for yield or some combination of both. Table 4-8 shows that this “lack of genetic diversity” can lead to improved genotypes that yield 329 kg ha<sup>-1</sup> lower than the global maximum. Considering that average genetic gains in soybean have been 15 kg ha<sup>-1</sup> yr<sup>-1</sup> (Boerma, 1979; Specht and Williams, 1984; Voldeng et al., 1997; Morrison et al., 1999; Wilcox et al., 1979), a

yield reduction of 300 kg ha<sup>-1</sup> relative to the maximum attainable is equivalent to 20 years of genetic gains through conventional plant breeding.

Given a large enough  $\Phi$  to allow the full exploration of the parameter space at the beginning of the maximization, variations in  $r_\phi$  are analogous to variation in selection pressure throughout the genetic improvement process. Table 4-8 shows that the lower the selection pressure, or higher  $r_\phi$ , the closer is the identified maximum yield to the global maximum. High selection pressure, as can be the case with a combination  $\Phi=1$  and  $r_\phi=0.01$ , can lead to rapid genetic gains; simulated yields increased from 1685 to 2891 kg ha<sup>-1</sup>. However, as it was postulated, high selection pressure led to yield stagnation; the maximum yield corresponding to  $\Phi=1$  and  $r_\phi=0.01$  was 349 kg ha<sup>-1</sup> lower than the global maximum.

Table 4-8. Maximum-yield sensitivity to variations in parameters  $\Phi$  and  $r_\phi$

$r_\phi$	$\Phi^*$		
	1	5	25
0.01	2891	3159	3167
0.25	2891	3214	3235
0.85	2948	3235	3240

\* Kirpatrick et al (1983) used temperature as a metaphor instead of  $\Phi$  due to its analogy with annealing of metals. In a plant breeding context  $\Phi$  is analogous to the breadth of the genetic base, and  $r_\phi$  is analogous to the selection pressure.

From this analysis we can conclude that both a high selection pressure and a narrow genetic base breadth can lead to yield stagnation. Although is difficult to make a generalization, our results suggest that the breadth of the genetic base is more important than the selection of an optimal selection pressure. Variations with  $\Phi$  for a given  $r_\phi$  are larger than variations with  $r_\phi$  for a given  $\Phi$ .

### Risks of Ignoring Epistatic and Pleiotropic Effects for Ideotype Design

Ignoring epistatic and pleiotropic effects for ideotype design in target environments led to overestimating potential genetic gains. To evaluate these effects we replace during the optimization process some genetic coefficients were estimated using functions of *E* loci (Table 3-4; Table 2-1). Simulated yields for ideotypes based on the optimization of *E* loci were 238 and 822 kg ha<sup>-1</sup> lower than those from ideotypes designed on the basis of genetic coefficients alone in Balcarce and Pergamino, respectively. Yield components did not vary relative to previous simulated ideotypes (Table 4-4). Seed weight was 236 mg in Balcarce simulations and 41 mg for Pergamino.

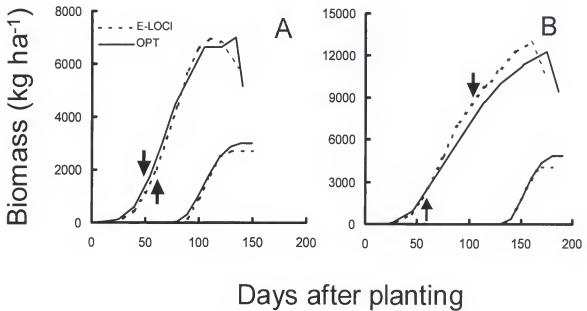


Figure 4-6. Simulated total and seed biomass for SA optimized genotypes grown in Balcarce (A) and Pergamino (B). OPT: optimization of genetic coefficients as continuous variables. *E*-LOCI indicates that optimization of genetic coefficients was done searching the *E*-LOCI space whenever possible. Arrows indicate the occurrence of R1.

For a given location, the differences in ideotype yields were caused by a reduction in the length of the crop cycle and by a relatively shorter seed fill duration (Fig. 4-6). However, the magnitude of these differences between ideotypes varied with location. The

genetic limitations derived by pleiotropic and epistatic effects were minimum in the simulated ideotype for Balcarce, and it included in its genotype dominant alleles at the loci *E4* and *E5*. Biomass accumulation curves are quite similar for both ideotypes (Fig. 4-6a). By only including dominant alleles for *E4* and *E5* the ideotype has low sensitivity to photoperiod, allowing the ideotype to adequately fit its growth cycle within the growing season. However, this is attained at the expense of a shorter seed fill duration.

In contrast, the ideotype for Pergamino included all loci from *E1* through *E4*. It was shown that these loci interact with each other and have pleiotropic effects on soybean development (Chapter 3), which is reflected in the dynamics of biomass accumulation and the timing of reproductive events (Fig. 4-6b). Both ideotypes strategy for yield maximization in Pergamino avoided the mid season drought (Fig. 4-2b) by delaying critical reproductive stages. As shown before, this strategy required the maximization of photoperiod sensitivity (Table 4-4). When the ideotype was designed by *E* loci optimization, CSDL and PPSEN were 12.39 h and 0.390 h<sup>-1</sup> respectively, which are close to the values found before for Pergamino (Table 4-4). Because *E* loci have pleiotropic effects, there was an unintended delay of time to flowering as shown by the increased photothermal time to flowering (24.8 vs. 15.5), and a shortening of the seed filling period as shown by a decreased in photothermal time between first seed and physiological maturity (31.0 vs.38.6).

Increasing yields in favorable years drove average yield maximization. Probability of exceedence curves shows that ideotypes outperformed probe genotypes throughout the range of climatic conditions in both Balcarce and Pergamino (Fig. 4-7). However, the main differences are striking towards the highest yields. This behavior is more evident in

ideotypes designed under the assumption of no pleiotropic and epistatic effects. The results presented here indicate that ignoring these effects can lead to overestimating genetic gains in some target environments, such as Pergamino in this case study.

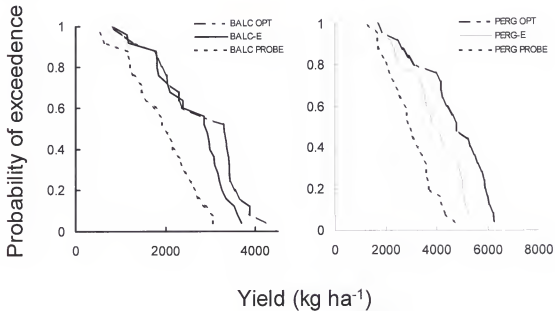


Figure 4-7. Probabilities of exceedence of yield simulated for probe, optimized genotypes and *E*-loci optimized genotypes in Balcarce (BALC) and Pergamino (PERG)

### Discussion and Conclusions

The search for ideotypes that maximize yield dates back to the sixties when Donald (1968) first proposed the concept of an ideal phenotype with certain physiological characteristics as a way to assist plant breeding. The concept was used then in the development of high-yield rice and wheat cultivars (Duvick, 2002) and has been extremely useful since then (Belford and Sedgley, 1991). Recently, breeders experimented with a new ideotype for wheat (Science, 1998) and rice (Cooper, 1999). However, the complexities of biological systems and quantitative traits make us question our abilities to identify plant characters and ways to enhance yield when more complex physiology and genetics than the one regulating dwarfism in wheat are involved. Recent

advances in molecular biology are helping us realize the magnitude and complexity of biological systems and help us characterize this complexity (The Arabidopsis Genome Initiative, 2000; Ideker et al., 2001; Staub and Serquen, 1996). Systems approaches are so far the best paradigm to study, understand, and manipulate complex systems. Although systems approaches were widely used to understand plant systems (Jones et al., 2003; Keating et al., 2003; van Ittersum et al., 2003), only recently have there been efforts to use molecular level knowledge to simulate plant traits (White and Hoogenboom, 1996; Reymond et al., 2003; Yin et al., 2003; Stewart et al., 2003; Chapter 3). We now have the opportunity to design plant ideotypes from very basic biological principles at the molecular level. Toward this end, we developed an approach for unsupervised ideotype design by linking a global optimization algorithm and a gene-based crop model.

### **Linking SA and CROPGRO-Soybean for Ideotype Design**

Crop models have been used for ideotype design (Boote and Tollenaar 1994; Boote et al. 2001; Boote et al., 2003; Aggarwal et al., 1997; White, 1998; Hunt, 1993; Kropff et al., 1995), and have also been linked to optimization algorithms (Hammer et al., 1996). These studies recognized that ideotype design requires optimizing multiple traits to attain significant increases in yield. Boote et al. (2001) showed that to attain a 10% increase in soybean yield there should be simultaneous changes in maximum photosynthesis, crop determinacy and seed-fill duration. Increases in soybean yield in our study for different environments in Argentina using simulated annealing were at least 40% relative to probe genotypes. Our results strengthen the concept that genetic improvement lays on the adequate selection of multiple traits for specific environments, and demonstrate that this approach can improve our abilities to enhance yields.

Breeding for multiple traits is challenging since the complexity of the yield response surface and the probability of finding local maxima increases with increasing number of traits. Studying the effects of genetic base breadth and selection pressure on genetic gains, we showed for Santa Rosa the existence of multiple local maximum (Table 4-8). This result confirmed our assumption and proves that previous approaches have limitations for ideotype design based on multiple trait optimizations. In contrast, our approach, based on simulated annealing, identified a global maximum in all cases (e.g., Fig. 4-3). Nonetheless, simulated annealing can produce suboptimal results if inadequate parameters are selected to guide the parameter search. The same analysis for Santa Rosa shows that this can be the case. Provided that adequate procedures for selected parameters  $\Phi$  and  $\Phi_n$  are followed (see Goffe et al., 1994), and test for convergence are conducted, we conclude that SA can improve ideotype design by minimizing the risk of identification of suboptimal genotypes.

CROPGRO-Soybean driven by SA was able to identify improved phenotypes for each target environment. As in previous studies, the ultimate combination of traits or genetic coefficients was identified (Hammer et al., 1996; Boote et al. 2001; Boote et al., 2003; Aggarwal et al., 1997; White, 1998; Hunt, 1993; Kropff et al., 1995). Because SA identified a global maximum, the pathway leading to maximum yield also provided additional information that could assist in plant breeding. The evolution of the combinations of genetic coefficients allows us to create a hierarchy of traits to determine an order for improvement. Results presented for Balcarce (Fig. 4-3) shows there were four phases of changes in the genetic coefficients. During the phase I, seed fill duration had the largest impact on yield increase provided that the photoperiod sensitivity did not

change. During the same phase, variations in traits such as seed weight or the juvenile phase were irrelevant for yield improvement. Only after seed fill duration was maximized did variations in specific leaf area have an impact on yield. Optimization results can provide the combination of traits corresponding to an ideotype, but also they can help guide the process that leads to it through parental selection.

### **Optimal Plant Traits for Target Environments**

Optimization of plant characters for yield maximization using a crop model identified traits for specific and broad adaptation, which were in agreement with those suggested from genetic, physiological and simulation studies. Seed fill duration was the single most important trait that increased yield in all target environments. Repeatable positive associations between soybean yield and estimates of seed filling durations were reported under wide range of environments (Hanway and Weber, 1971; Egli and Leggett, 1973; Boote, 1981; Smith and Nelson, 1986; Hanson, 1985). Analysis of past genetic improvement in soybeans shows that higher yields of newer cultivars were associated with a longer seed-filling period (Gay et al., 1980). Boote et al. (2001) used simulation to study past genetic improvement and ideotype design and arrived to the same conclusion.

Increasing photoperiod sensitivity during post-flowering further extended the seed filling period. Our results showed consistent increases in R1PPO and reductions in CSDL relative to probe genotypes across target environments. Increased photoperiod sensitivity in post-flowering also slowed seed growth rate (Thomas and Raper, Jr., 1976) hence increasing soybean yield potential (Hanson and Burton, 1994). Genotypes with slower seed growth would have a reduced nitrogen demand and remobilization from leaf tissue. These physiological changes would extend leaf area duration and maintain higher photosynthetic rates. Buttery et al. (1981) showed a strong relationship between specific

leaf nitrogen and leaf photosynthesis, and between photosynthesis during seed fill and yield. The maintenance of green leaf area, which can be associated with stay-green trait, increases canopy photosynthesis and yields, provided that the protein concentration in the seed remains constant. Modern high yielding genotypes have reduced rate of leaf senescence and increased LAI during seed-fill (Kumudini et al., 2001).

Reductions in the pre-flowering or early post-flowering duration compensated for the increases in seed fill. For all target environments, ideotypes had a reduced sensitivity to photoperiod during the vegetative phase. Soybean breeders have actively searched for loci controlling photoperiod insensitivity (Tasma and Shoemaker, 2003; Tasma et al., 2001; Jun Abe et al., 2003). One of the mechanisms involved in this study incorporated a long juvenile phase, which extended the time between emergence and flowering during which the plant is not sensitive to the photoperiodic stimuli. This trait is rare among cultivars of early maturity groups typically grown in the Pampas or the USA. However, recent evidence suggests that including the long juvenile trait in maturity groups IV and V can increase yields at early plantings (Tomkins and Shipe, 1996).

Our simulation results suggest that soybean ideotypes would have low specific leaf area, hence high maximum leaf photosynthesis. Genetic evidence supports this proposition. It was shown that canopy photosynthesis during the reproductive period was correlated with seed yield in a diverse group of genotypes including plant introductions and improved cultivars (Wells et al., 1982; Boerma and Ashley, 1988). Morrison et al. (1999) showed that yield and leaf photosynthesis increased 0.5% per year since 1930 in response to selection in short season cultivars. Furthermore, the increased photosynthesis was associated with a simultaneous reduction in specific leaf area as suggested by our

simulation results. Previous simulation studies also showed a positive relationship between increases in leaf photosynthesis, canopy photosynthesis and yield (Boote and Tollenaar, 1994). However, other researchers found low to nil associations suggesting that selection for higher leaf photosynthesis is futile (Thompson et al., 1995; Kumudini, 2002).

These conflicting results can be explained by inconsistencies between experiments in the timing between the measurements, the environment and phenological stage (Kumudini, 2002). An alternative hypothesis can be proposed based on the negative association between seed yield and oil content with respect to protein concentration (Chung et al., 2003). Cultivars with low seed protein content would require less nitrogen remobilization from leaves, which can maintain higher photosynthetic rates and for a longer period of time supporting higher seed yield and oil content. Thompson et al. (1995) evaluated the relationship between leaf carbon exchange rate and yield using F6 populations derived from crosses between A3127 and Elgin (both with 37% protein) and eight plant introductions with protein content varying between 43 and 46%. The lack of an association between photosynthesis and yield may be explained by confounding effects introduced by the segregation of protein content in their plant material. Furthermore, it is apparent that these plants were sink-limited by water stress in 1991 and low radiation and cool temperatures in 1992 (Thompson et al., 1995). Note that for canopy or maximum leaf photosynthesis to have an impact on yield there must be sink limitation. Simulations for Balcarce showed that variations in specific leaf area, hence in leaf photosynthesis (Dornhoff and Shibles, 1970), had an impact on yield (Fig. 4-3, Fig.

4-4), but only after seed fill duration was maximized. Boerma and Ashley (1988) showed that cultivars with high photosynthesis and yield also had longer seed-fill duration.

Traits conferring soybean specific adaptation were targeted to avoid or minimize the impact of water stress during the reproductive period. Under terminal drought, ideotypes with high seed weight improved soybean adaptation. Seed weight is associated with early vigor allowing the plant to increase leaf area and light interception early in the season assuring setting a reduced number of seeds and the accumulation of N for future remobilization. Because seed growth is less affected than seed number under light or water stress conditions (Frederick et al., 1991; Egli, 1998; Jiang and Egli, 1993), allocating assimilates to seed growth during the terminal drought is more beneficial than its allocation to more vulnerable structures such as small pods.

In contrast, the strategies for mid season drought in Santa Rosa and Pergamino were based on stress avoidance. Pod addition and reproductive stages were delayed in the growing season to take advantage of early Fall rainfall. Extension of pod addition duration will also create a resilient crop by spreading in time the determination of seed number. In chapter 2 we suggested that increasing pod addition duration could increase pod number and yields. Kantolic and Slafer (2001) proposed a similar hypothesis based on experimental results near Pergamino. The determination of higher number of seeds led to a reduction in seed weight provided that the ideotype is source and not sink limited.

Under more severe water stress conditions during the 50's in Santa Rosa, the optimum strategy was shortening the crop cycle duration, the duration of canopy expansion and leaf size, resulted in the reduction of leaf area, which in turn delayed water use for later more critical stages. This result illustrates tradeoff between carbon

assimilation and water use. It also shows that optimal ideotypes can vary with time if climate varies, raising the question of what period to select for yield maximization without incurring additional risks for underestimating water stress limitations or sub-utilization of available resources. Considering that a breeding program takes about 15 years from the beginning of the program to the release of the variety, this example for Santa Rosa shows that the characters for which the cultivar was selected for could be inadequate at the moment the cultivar is to be grown. By designing ideotypes in silico, selecting by molecular markers, the time span between conception of the program and release of the variety could be reduced.

We showed that there is not a unique strategy to increase yield based on selection for yield components; both increases in seed weight and seed number increased yields depending on the environment. Based on our simulation results, we conclude that increasing seed number (as found by Kantolic and Slafer (2001) and as we suggested in chapter 2) may be adequate to increase yield for temperate climates in mid latitudes such as in Pergamino, but this strategy may not be valid for other environments. Other strategies may be more adequate under different environmental challenges. Crop models linked to global optimization algorithms can help make informed decision about traits and strategies that can increase yield in diverse environments.

### **Genetic Base Breadth and Crop Yields**

During domestication of crops and plant breeding, genetic diversity has been segregated and maintained in populations with relatively narrow genetic basis (Loomis and Connor, 1992). There is consensus that a narrow genetic base can limit soybean yield gains in the US (Kisha et al., 1998; Manjarez-Sandoval et al., 1997). Its effects on potential yield gains were studied through the effects of the coefficient of parentage on

the genetic variance (Manjarrez-Sandoval et al., 1997). Only populations derived from crosses with low coefficient of parentage had predicted high genetic gains. This is a necessary condition, but not sufficient for guarantying success. A low coefficient of parentage would improve genetic gains in yield only if divergence at the molecular level reflects variability in traits contributing to yield in that given environment. Our analysis of the evolution of genetic coefficients during the maximization process demonstrates that contribution of genetic coefficients to yield maximization is relative to other coefficients. Considering phase I of the optimization process, a cross between two populations varying in all genes but those regulating seed fill duration would have a low coefficient of parentage, but its relationship with genetic variance in yield and associated genetic gains would be from low to nil.

However, our results for Santa Rosa support this conclusion and show it is of general validity. Yields of ideotypes using low values of  $\Phi$ , which would correspond to a high coefficient of parentage, were about 300 kg ha<sup>-1</sup> lower than the global maximum yield, which can be only identified from a broad genetic base. Because we used a simulation approach, any measure of coefficient of parentage is related to the breadth in all traits contributing to yield maximization by definition. In addition the global maximum yield is known, allowing the estimation of yield losses and existence of local optima. We conclude that narrow genetic base reduces genetic gains due to convergence to local maximum.

### **Gene-based Models for Ideotype Design**

Impacts of crop models in plant improvement have not met our expectations despite the great potential they have to offer (White, 1998). Indeed, most applications of crop models as a tool to assist plant breeding were conducted by modelers rather than

breeders. We can propose at least four reasons that may hamper the use of models in “real” plant breeding: 1) breeder’s lack of expertise in crop modeling, 2) reliance on field trials to fit model coefficients (White, 1998), 3) inadequate links between genes and plant traits, and 4) the inadequate representation of epistatic and pleiotropic effects.

The inadequate representation of epistasis and pleiotropic effects creates a risk of overestimating genetic gains by selecting infeasible combinations of traits. Genetic gains for Pergamino were overestimated when unlinked genetic coefficients were optimized compared to the optimization of *E* loci. Overestimation of genetic gains for Balcarce simulations was less important. These results question the validity of previous attempts of ideotype design using simulation models without parameters linked to loci (e.g., Hammer et al., 1996; Hunt, 1993; Kropff et al., 1995). There was no evidence to determine the magnitude of the linkage problem in previous studies.

By searching the *E* loci space we found suboptimal genotypes that are more realistic as ideotypes in the short term. However, higher yields are possible provided that different regulatory mechanisms of plant development are incorporated in the plant. We need to understand better the genetic controls underlying the regulation of development to bypass the limitation currently imposed by the regulatory mechanisms associated with *E* loci.

By linking a gene-based model with a global optimization algorithm, the limitations to the use of crop models in plant breeding are in part removed. The gene-based model driven by SA was successful in identifying global maximum yields. Since the model is parameterized with *E* loci information, breeders would not need to rely solely on field experimentation to estimate parameters, but on more time-efficient and

familiar techniques based on molecular markers. There is potential to fully parameterized crop models with available information about makers and plant traits. Table 4-2 summarizes potential markers that can ultimately be used to replace existing genetic coefficients in CROPGRO-Soybean provided adequate experimentation is conducted to derive the relationships between genetic coefficients and marker loci. Driving the crop model by selecting loci instead of genetic coefficients, epistatic and pleiotropic effects are better represented. In addition, the search space reduces significantly since the SA algorithm searches a discrete rather than a continuous space, for which it was originally designed (Kirkpatrick et al., 1983), reducing the algorithm run time. These modifications relative to more conventional approaches for ideotype design can improve the interface between modeling and breeding, thus helping realize the potential of crop models to assist genetic improvement.

## CHAPTER 5 SUMMARY AND CONCLUSIONS

The research described in this dissertation contributes to the development of the emerging discipline of systems biology, with emphasis on the simulation of plant growth and development for ideotype and food production systems design. The overall objective of this work was to develop and test a systems approach for ideotype design based on previously characterized alleles at selected loci. To achieve this objective, in Chapter 2 I characterized the alleles at 7 soybean loci, which regulate growth habit and responses to photoperiod. In Chapter 3 I used the knowledge gained and the experimental data set to replace parameters controlling development in CROPGRO-Soybean with linear regression functions of *E* loci. I developed a new gene-based biophysical model for soybean. The model was validated with independent data sets including variety trials. Finally in Chapter 4 I tailored the gene-based model with a global optimizer to design ideotypes for target environments.

Chapter 2 used soybean as a model organism to study the genetic control of response to photoperiod mediated by *dt* and *E* loci during the reproductive period, and to evaluate the effects of these loci on fruit number. Previous research reported the effects of *E* loci on time to flowering and maturity. However, we had incomplete knowledge about the effects of *E* loci on critical phases of the reproductive development of soybean. A field experiment was conducted to test the hypotheses:

- The *dt* and *E* loci regulate the duration of the following periods: a) from first flower to first pod; b) pod addition; c) seed filling; and d) from first flower to the onset of seed development.

- *E* loci regulate pod number by affecting the rate of pod addition.
- *E* loci regulate duration of pod addition by regulating the onset of seed development.

This Chapter showed that the *dt* locus and *E* loci regulate the photoperiodic response of the duration from first flower to first pod (Fig.2-5), the duration of the critical period of pod addition (Table 2-3, Fig.2-4), time to R5 as the estimator for the onset of seed filling (Table 2-4) and the seed-filling duration as estimated by the duration between R5 and R7 (Table 2-5).

This research showed that *E* loci regulate pod addition duration through the regulation of the time to the onset of seed growth, as shown by the association between pod addition duration and time to R5 (Fig.2-6). Finally, *E* and *dt* loci controlled fruit number by regulating pod addition duration. The results obtained do not support conclusively the relationship between rate of pod addition and pod number.

In Chapter 3 I developed and evaluated a gene-based biophysical model that simulates soybean growth and development using experimental data generated in Chapter 2. This was done by incorporating relationships between *E* loci and model parameters into the physiological model CROPGRO-Soybean. This constitutes a step forward with respect to previous models to predict time to flowering from *E* loci information. In contrast to modeling processes alone, the integration of physiological processes in CROPGRO-Soybean allows one to study the effects of genes controlling development of other physiological processes and traits of agronomic interest, (Fig.3- 6). CROPGRO-Soybean predicted accurately time to flowering and post-flowering development phases (Fig.3- 3; Table 3-3; Table 3-5). The prediction skill showed by the new gene-based model was comparable with that of Genegro for dry bean (White and

Hoogenboom, 1996) and CROPGRO-Soybean as a stand-alone model. We showed that CROPGRO-Soybean could accurately predict pod number (Fig.3- 4) by adequately simulating pod addition duration (Fig.3- 3). Furthermore, the model predicted correctly relationships between physiological processes, such as the association between pod addition duration and the time to seed growth (Fig.3- 5).

For the first time, a gene-based model was tested for its ability to reproduce yields and development at the field scale by only knowing the genetic makeup of the cultivar. Because prediction errors using a gene-based approach are comparable with conventional parameter estimation, gene-based models are a practical alternative for yield simulation. The gene-based model predicted 75% of the variance in time to maturity and 54% of the yield variance in variety trials conducted in Illinois. Genetic-based approaches can decrease the requirements for expensive and time-consuming experimentation for model parameterization. Failure to simulate yield and development for the Savoy cultivar shows that there is potential for further reducing uncertainties, errors and risks involved in the development and implementation of gene-based approaches.

Chapter 4 describes a methodology for ideotype design for target environments that coupled crop simulation models and a global optimization algorithm (simulated annealing). I introduced a new metaphor for this optimization approach, based on the equivalence between simulated annealing cooling parameters, and selection pressure and genetic base breadth. The coupled model identified ideotypes yielding at least 40% more than actual varieties grown in Argentina (Table 4-4). These results strengthen the concept that genetic improvement lies in the adequate selection of multiple traits, and demonstrate

that our approach could improve our abilities to enhance yields by avoiding local maxima (Table 4-8; Fig.4-3).

I showed that the inadequate representation of epistasis and pleiotropic effects of genes on physiological traits increases the risk of overestimating genetic gains by selecting unfeasible combinations of traits. Genetic gains for Pergamino were overestimated when genetic coefficients were optimized to maximize yields relative to the optimization of *E* loci. This result questions the validity of previous attempts of ideotype design using simulation models without parameters linked to loci or gene action (e.g., Hammer et al., 1996; Hunt, 1993; Kropff et al., 1995). There was no evidence to determine the magnitude of the linkage problem in previous studies, but the present work shows the advantage of gene-based approaches for ideotype design.

With ideotypes identified by the application of this method we do not claim to provide the ultimate recipe for the breeder. Instead, we seek to identify neighborhoods of traits or gene combinations around which plant breeders should focus their efforts using more traditional approaches. By designing ideotypes in silico, and selecting with molecular markers, the time span between conception of the program and release of a variety could be reduced.

I have high expectations for the applicability of this approach for developing cultivars adapted to specific environments by exploiting favorable gene-by-environment interactions. This implies a change in the commercial plant breeding paradigm that seeks to improve yields based on traits that confer broad adaptation. Poor farmers live in very diverse environments and require specific solutions and cultivars. With the recent revitalization of the CGIAR system (Kennedy, 2003) we can envision a system that

produces and characterizes their large germplasm collections at all loci. This information can be used to drive gene-based models to design cultivars best adapted to local conditions.

## LIST OF REFERENCES

- AACREA. 1997. Soybean Technical bulletin No.58 (In Spanish). Realizacion Grafica Integral, Buenos Aires, Argentina
- Acocck, B., and J.F. Reynolds. 1997. Modularity in plant models. *Ecol. Model.* 94:1-6
- Aggarwal, P.K., M.J. Kropff, K.G. Cassman, and H.F.M. ten Berge. 1997. Simulating genotypic strategies for increasing rice yield potential in irrigated, tropical environments. *Field Crops Res.* 51:5-17
- Ahuja, R.K., T.L. Magnanti, and J.B. Orlin. 1993. Network flows. Theory, algorithms, and applications. Prentice Hall, Upper Saddle River, NJ.
- Akaike, H. 1974. A new look at statistical model identification. *IEEE Trans. Automat. Contr.* AU-19:716-722
- Akkaya, M.S., A.A. Bhagwat, and P.B. Cregan. 1992. Length polymorphisms of simple sequence repeat DNA in soybean. *Genetics* 132:1131-1139
- Alter, O., P.O. Brown, and D. Botstein. 2000. Singular value decomposition for genome-wide expression data processing and modeling. *Proc. Natl. Acad. Sci. U.S.A.* 97:10101-10106
- Allard, R.W., and A.D. Bradshaw. 1964. Implications of genotype-environmental interactions in applied plant breeding. *Crop Sci.* 4:503-508
- Allen, F., G. Almasi, W. Andreoni, D. Beece, B.J. Berne, A. Bright, J. Brunheroto, C. Cascaval, J. Castanos, P. Coteus, P. Crumley, A. Curioni, M. Denneau, W. Donath, M. Eleftheriou, B. Fitch, B. Fleischer, C.J. Georgiou, R. Germain, M. Giampapa, D. Gresh, M. Gupta, R. Haring, H. Ho, P. Hochschild, S. Hummel, T. Jonas, D. News, M. Newton, R. Philhower, T. Picunko, J. Pitera, M. Pitman, R. Rand, A. Royyuru, V. Salapura, A. Sanomiya, R. Shah, Y. Sham, S. Singh, M. Snir, F. Suits, R. Swetz, W.C. Swope, N. Vishnumurthy, T.J.C. Ward, H. Warren, and R. Zhou. 2001. *IBM Systems Journal* 40:310-327
- Andrade, F.H., P. Calvino, A. Cirilo, and P. Barbieri. 2002. Yield responses to narrow rows depend on increased radiation interception. *Agron. J.* 94:975-980
- The Arabidopsis Genome Initiative.* 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature (London)* 408:796-815

- Aumadu, H., R.J. Summerfield, R.H. Ellis, and A. Qi. 1998. Variation in the durations of the photoperiod-sensitive and photoperiod-insensitive phases of post-first flowering development in maturity isolines of soybean [*Glycine max* (L.) Merrill]'Clark'. *Ann. Bot.* 82:773-778
- Aukerman, M.J., and H. Sakai. 2003. Regulation of flowering time and floral organ identity by microRNA and its *APETALA2*-Like target genes. *The Plant Cell* 15:2730-2741
- Belford, R.K., and R.H. Sedgley. 1991. Conclusions: Ideotypes and physiology: Tailoring plants for increased production. *Field Crops Res.* 26:221-226
- Bernard, R.L. 1971. Two major genes for time of flowering and maturity in soybeans. *Crop Sci.* 11:242-244
- Bhalla, U.S., and R. Iyengar. 1999. Emergent properties of networks of biological signaling pathways. *Science* (Washington DC) 283:381-387
- Board, J.E., and J.R. Settimi. 1986. Photoperiod effect before and after flowering on branch development in determinate soybean. *Agron. J.* 78:995-1002
- Board, J.E., and J.R. Settimi. 1988. Photoperiod requirements for flowering and flower production in soybean. *Agron. J.* 80:518-525
- Boerma, H.R. 1979. Comparison of past and recently developed soybean cultivars in maturity groups VI, VII and VIII. *Crop Sci.* 19:611-613
- Boerma, H.R., and D.A. Ashley. 1988. Canopy photosynthesis and seed fill duration in recently developed soybean cultivars and selected plant introductions. *Crop Sci.* 28:137-140
- Boote, K.J. 1981. Response of soybeans in different maturity groups to march planting in southern USA. *Agron. J.* 73:854-859
- Boote, K.J., J.W. Jones, W.D. Batchelor, E.D. Nazfiger, and O. Myers. 2003. Genetic coefficients in the CROPGRO-Soybean model: links to field performance and genomics. *Agron. J.* 95:32-51
- Boote, K.J., J.W. Jones, and G.H. Hoogenboom. 1998. Simulation of crop growth: CROPGRO Model. p.651-693. *In* R.M. Peart and R.B. Curry (ed.) *Agricultural systems modeling and simulation*. Marcel Dekker, New York
- Boote, K.J., M.J. Kroft, and P.S. Brindaban. 2001. Physiology and modeling of traits in crop plants: implications for genetic improvement. *Agric. Syst.* 70: 395-420
- Boote, K.J., and M. Tollenaar. 1994. Modeling genetic yield potential. p. 533-561. *In* Boote, K.J. et al. (ed.) *Physiology and determination of crop yield*. ASA, CSSA and SSSA, Madison, WI

- Bradford, K.J., and A.J. Trewavas. 1994. Sensitivity thresholds and variable time scales in plant hormone action. *Plant Physiol.* 105:1029-1036
- Braun, P., and J. LaBaer. 2003. High throughput protein production for functional proteomics. *TRENDS Biotechnol.* 21:383-388
- Brazhnik, P., A. de la Fuente, and P. Mendes. 2002. Gene networks: how to put the function in genomics. *TRENDS Biotechnol.* 20:467-472
- Brown, P.O., and D. Botstein. 1999. Exploring the new world of the genome with DNA microarrays. *Nature Genet.* 21:33-37
- Buchanan, B.B., W. Gruissem, and R.L. Jones. 2000. Biochemistry and molecular biology of plants. ASPP, Waldorf, MD
- Buttery, B.R., R.I. Buzzell, and W.I. Findlay. 1981. Relationships among photosynthetic rate, bean yield and other characters in field-grown cultivars of soybean. *Can. J. Plant Sci.* 61:191-198
- Carlson, J.B. 1973. Morphology. p.17-95. *In* Soybeans: Caldwell (ed.) Improvement, production, and uses. Agron. Monogr. 16. ASA, Madison, WI.
- Chapman, S., M. Cooper, D. Podlich, and G. Hammer. 2003. Evaluating plant breeding strategies by simulating gene action in dryland environment effects. *Agron. J.* 95:99-113
- Charles-Edwards, D.A. 1984a. On the ordered development of plants. 1. An hypothesis. *Ann. Bot.* 53:699-707
- Charles-Edwards, D.A. 1984b. On the ordered development of plants. 2. Self-thinning in plant communities. *Ann. Bot.* 53:709-714
- Charles-Edwards, D.A., and D.F. Beech. 1984. On the ordered development of plants. 3. Branching by the legume *Cyamopsis tetragonoloba* (Guar). *Ann. Bot.* 54:673-679
- Chung, J., H.L. Babka, G.L. Graef, P.E. Staswick, D.J. Lee, P.B. Cregan, R.C. Schoemaker, and J.E. Specht. 2003. The seed protein, oil, and yield QTL on soybean linkage group I. *Crop Sci.* 43:1053-1067
- Cober, E.R., D.W. Stewart, and H.D. Voldeng. 2001. Photoperiod and temperature responses in early-maturing near-isogenic soybean lines. *Crop Sci.* 41:721-727
- Cober, E.R., J.W. Tanner, and H.D. Voldeng. 1996. Genetic control of photoperiod response in early-maturing, near-isogenic soybean lines. *Crop Sci.* 36:601-605
- Cober, E.R., and H.D. Voldeng. 2001. A new soybean maturity and photoperiod-sensitivity locus linked to *E1* and *T*. *Crop Sci.* 41:698-701

- Colson, J., A. Bouniols, and J.W. Jones. 1995a. Soybean reproductive development: adapting a model for European cultivars. *Agron. J.* 87:1129-1139
- Colson, J., D. Wallach, A. Bouniols, J.B. Denis, and J.W. Jones. 1995b. Mean squared error of yield prediction by SOYGR0. *Agron J.* 87:397-402
- Cooper, M. 1999. Concepts and strategies for plant adaptation research in rainfed lowland rice. *Field Crops Res.* 64:13-34
- Cooper, M., S.C. Chapman, D.W. Podlich, and G.L. Hammer. 2002. The GP problem: Quantifying gene-to-phenotype relationships. In *Silico Biology* (Available on-line at <http://www.bioinfo.de/isb/>; verified 10 Oct. 2003)
- Corana, A., M. Marchesi, C. Martini, S. Ridella. 1987. Minimizing multimodal functions of continuous variables with the simulated annealing algorithm. *ACM Trans. Math. Softw.* 13:262-280
- Cregan, P.B., T. Jarvik, A.L. Bush, R.C. Schoemaker, K.G. Lark, A.L. Kahler, N. Kaya, T.T. vanToai, D.G. Lohnes, L. Chung, J.E. Specht. 1999. *Crop Sci.* 39:1464-1490
- Csete, M.E., and J.C. Doyle. 2002. Reverse engineering of biological complexity. *Science* (Washington DC) 295:1664-1669
- Cure, J.D., R.P. Patterson, C.D. Raper, Jr., and W.A. Jackson. 1982. Assimilate distribution in soybeans as affected by photoperiod during seed development. *Crop Sci.* 22:1245-1250
- Curtis, D.F., J.W. Tanner, B.M. Luzzi, and D.J. Hume. 2000. Agronomic and phenological differences of soybean isolines differing in maturity and growth habit. *Crop Sci.* 40:1624-1629
- Daniell, H., and A. Dhingra. 2002. Multigene engineering: dawn of an exciting new era in biotechnology. *Curr. Opin. Biotechnol.* 13:136-141
- Davidson, E.H., J.P. Rast, P. Oliveri, A. Ransick, C. Caletani, C.H. Yuh, T. Minokawa, G. Amore, V. Hinman, C. Arenas-Mena, O. Otim, C.T. Brown, C.B. Livi, P.Y. Lee, R. Revilla, A. Rust, Z. Pan, M.J. Schilstra, P.J.C. Clarke, M.I. Arnone, L. Rowen, R.A. Cameron, D.R. McClay, L. Hood, H. Bolouri. 2002. A Genomic Regulatory Network for Development. *Science* (Washington DC) 295:1669-1678
- de Wit, C.T., and F.W.T. Penning de Vries. 1983. Crop growth models without hormones. *Neth. J. Agric. Sci.* 31:313-323
- Dornhoff, G.M., and R.M. Shibles. 1970. Varietal differences in net photosynthesis of soybean leaves. *Crop Sci.* 10:42-45
- Duggan, D.J., M. Bittner, Y. Chen, P. Meltzer, and J.M. Trent. 1999. Expression profiling using cDNA microarrays. *Nature Genet.* 21:10-14

- Duvick, D.N. 2002. Theory, empiricism and intuition in professional plant breeding. *In* D.A. Cleveland and D. Soleri (ed.) *Farmers, scientists and plant breeding*. CAB International, New York
- Egli, D.B. 1998. Seed biology and the yield of grain crops. CAB International, New York
- Egli, D.B., and W.P. Bruening. 2000. Potential of early-maturing soybean cultivars in late plantings. *Agron. J.* 92:532-537
- Egli, D.B., R.D. Guffy, and J.E. Leggett. 1985. Partitioning of assimilate between vegetative and reproductive growth in soybeans. *Agron. J.* 77:917-922
- Egli, D.B., and J.E. Leggett. 1973. Dry matter accumulation pattern in determinate soybeans. *Crop Sci.* 13:220-222
- El-Assal, S.E., C. Alonso-Blanco, A.J.M. Peeters, V. Raz, and M. Koornneef. 2001. A QTL for flowering time in *Arabidopsis* reveals a novel allele of *CRY2*. *Nature Genet.* 29:435-440
- Elizondo, D.A., R.W. McClendon, and G. Hoogenboom. 1994. Neural network models for predicting flowering and physiological maturity of soybean. *Trans. ASAE* 37:981-988
- Ellis, R.H., H. Asumadu, A. Qi., and R.J. Summerfield. 2000. Effects of photoperiod and maturity genes on plant growth, partitioning, radiation use efficiency, and yield in soybean [*Glycine max* (L.) Merrill] 'Clark'. *Ann. Bot.* 85:335-343
- Fehr, W.R., and C.E. Caviness. 1977. Stages of soybean development. Spec. Rep. 80. Iowa Agric. Home Econ. Exp. Stn., Iowa State Univ. Ames
- Ferreira, R.A., H.P. Apezteguia, R. Sereno, and J.W. Jones. 2002. Reduction of soil water spatial sampling density using scaled semivariograms and simulated annealing. *Geoderma* 110:265-289
- Fisher, J.E. 1963. The effects of short days on fruitset as distinct from flower formation in soybeans. *Can. J. Bot.* 41:871-873
- Fraser, J., D.B. Egli, J.E. Leggett. 1982. Pod and seed development in soybean cultivars with differences in seed size. *Agron. J.* 74:81-85
- Frederick, J.R., J.T. Woolley, J.D. Hesketh, and D.B. Peters. 1991. Seed yield and agronomic traits of old and modern soybean cultivars under irrigation and soil water-deficit. *Field Crops Res.* 27:71-82
- Friedman, N. 2003. Probabilistic models for identifying regulation networks. *Bioinformatics*, 19:57

- Gay, S., D.B. Egli, and D.A. Reicosky. 1980. Physiological aspects of yield improvement in soybeans. *Agron. J.* 72:387-391
- Ghaemmaghani, S., W. Huh, K. Bower, R. Howson, A. Belle, N. Dephoure, E.K. O'Shea, and J.S. Weissman. 2003. Global analysis of protein expression in yeast. *Nature (London)* 425:737-741
- Gifford, R.M., J.H. Thorne, W.D. Hitz, and R.T. Giaquinta. 1984. Crop productivity and photoassimilate partitioning. *Science (Washington DC)* 225:801-808
- Gilman, A., and A.P. Arkin. 2002. Genetic "code": Representations and dynamical models of genetic components and networks. *Annu. Rev. Genomics Hum. Genet.* 3:341-369
- Gizlice, Z., T.E. Carter, Jr., and J.W. Burton. 1994. Genetic base for North American public soybean cultivars released between 1947 and 1988. *Crop Sci.* 34:1143-1151
- Goff, S.A., D. Ricke, T. Lan, G. Presting, R. Wang, M. Dunn, J. Glazebrook, A. Sessions, P. Oeller, H. Varma, D. Hadley, D. Hutchinson, C. Martin, F. Katagiri, B. Markus Lange, T. Moughamer, Y. Xia, P. Budworth, J. Zhong, T. Miguel, U. Paszkowski, S. Zhang, M. Colbert, W. Sun, L. Chen, B. Cooper, S. Park, T.C. Wood, L. Mao, P. Quail, R. Wing, R. Dean, Y. Yu, A. Zharkikh, R. Shen, S. Sahasrabudhe, A. Thomas, R. Cannings, A. Gutin, D. Pruss, J. Reid, S. Tavtigian, J. Mitchell, G. Eldredge, T. Scholl, R.M. Miller, S. Bhatnagar, N. Adey, T. Rubbano, N. Tusneem, R. Robinson, J. Feldhaus, T. Macalma, A. Oliphant, and S. Briggs. 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science (Washington DC)* 296:92-100
- Goffé, W.L., G.D. Ferrier, and J. Rogers. 1994. Global optimization of statistical functions with simulated annealing. *J. Econom.* 60:65-99
- Gonze, D., J. Halloy, and A. Goldbeter. 2002. Robustness of circadian rhythms with respect to molecular noise. *Proc. Natl. Acad. Sci. U.S.A.* 99:673-678
- Grimm, S.S., J.W. Jones, K.J. Boote, and J.D. Hesketh. 1993. Parameter estimation for predicting flowering date of soybean cultivars. *Crop Sci.* 33: 137-144
- Grimm, S.S., J.W. Jones, K.J. Boote, and D.E. Herzog. 1994. Modeling the occurrence of reproductive stages after flowering for 4 soybean cultivars. *Agron. J.* 86:31-38
- Guffy, R.D., J.D. Hesketh, R.L. Nelson, and R.L. Bernard. 1991. Seed growth rate, growth duration and yield in soybean. *Biotronics* 20:19-30
- Hadley, P., E.H. Roberts, R.J. Summerfield, and F.R. Minchin. 1984. Effects of temperature and photoperiod on flowering in soya bean [*Glycine max* (L.) Merrill]: a quantitative model. *Ann. Bot.* 53:669-681

- Hall, A.J., C.M. Rebella, C.M. Ghersa, and J.P. Culot. 1992. Field crop systems of the Pampas. p. 413-450. *In* Pearson, C.J. (ed.) Field crop ecosystems of the world. Elsevier, Exeter, UK
- Hammer, G.L., D.G. Butler, R.C. Muchow, and H. Meinke. 1996. Integrating physiological understanding and plant breeding via crop modeling and optimization. p.419-442. *In* Cooper, M. and G.L. Hammer (ed.) Plant adaptation and genetic improvement. CAB International, New York
- Hanson, W.D. 1985. Association of seed yield with partitioned lengths of the reproductive period in soybean genotypes. *Crop Sci.* 25:525-529
- Hanson, W.D., and J.W. Burton. 1994. Control of rate of seed development and seed yield potential in soybean. *Crop Sci.* 34:131-134
- Hanway, J.J., and C.R. Weber. 1971. Dry matter accumulation in eight soybean (*Glycine max* L. Merrill) varieties. *Agron. J.* 63:227-230
- Hart, R.P.S., M.T. Larcombe, R.A. Sherlock, and L.A. Smith. 1998. Optimization techniques for a computer simulation of a pastoral dairy farm. *Computers and Electronics in Agriculture* 19:129-15
- Hartwell, L.H., J.J. Hopfield, S. Leibler, and A.W. Murray. 1999. From molecular to modular cell biology. *Nature (London)* 402:C47-C52
- Heindl, J.C., and W.A. Brun. 1983. Light and shade effects on abscission and C-14 photoassimilate partitioning among reproductive structures in soybean. *Plant Physiol.* 73:434-439
- Hochberg, Y., and A.C. Tamhane. 1987. Multiple Comparison Procedures. Wiley, New York
- Holter, N.S., A. Maritan, M. Cieplak, N. Federoff, and J.R. Banavar. 2001. Dynamic modeling of gene expression data. *Proc. Natl. Acad. Sci. U.S.A.* 98:1693-1698
- Hoogenboom, G., and J.W. White. 2003. Improving physiological assumptions of simulation models by using gene-based approaches. *Agron. J.* 95:82-89
- Hoogenboom, G., J.W. White, J. Acosta-Gallegos, R.G. Gaudiel, J.R. Myers, and M.J. Silbernagel. 1997. Evaluation of a crop simulation model that incorporates gene action. *Agron. J.* 89:613-620
- Huang, E., S. Ishida, J. Pittman, H. Dressman, A. Bild, M. Kloos, M. D'Amico, R.G. Pestell, M. West, and J.R. Nevins. 2003. *Nature Genet.* 34:226-166
- Huang, J., C. Pray, and S. Rozelle. 2002. Enhancing the crops to feed the poor. *Nature (London)* 418:678-684

- Hunt, L.A. 1993. Designing improved plant types: a breeder's viewpoint. p. 3-18. *In* F. Penning de Vries et al. (ed.) *Systems approaches for agricultural development* Kluwer Academic Publishers, Dordrecht
- Hunt, L.A., and K.J. Boote. 1998. Data for model operation, calibration and evaluation. p. 41-54. *In* Tsuji, G.Y. et al. (ed.) *Understanding options for agricultural production*. Kluwer Academic Publishers, Dordrecht
- Hunt, L.A., S. Pararajasingham, J.W. Jones, G. Hoogenboom, D.T. Imamura, R.M. Ogoshi. 1993. GENCALC: Software to facilitate the use of crop models for analyzing field experiments. *Agron. J.* 85:1090-1094
- Ideker, T., T. Galitski, and L. Hood. 2001. A new approach to decoding life: systems biology. *Annu. Rev. Genomics Hum. Genet.* 2:343-372
- Jiang, H., and D.B. Egli. 1993. Shade induced changes in flower and pod number and flower and pod abscission in soybean. *Agron. J.*, 85:221-225
- Johnson, H.W., H.A. Borthwick, and R.C. Leffel. 1960. Effects of photoperiod and time of planting on rates of development of the soybean in various stages of the life cycle. *Bot. Gaz.* 122:77-95
- Jones, J.W. 1998. Model integration and simulation tools. *Acta Hort.* 456:411-417
- Jones, J.W., K.J. Boote, G. Hoogenboom, S.S. Jagtap, and G.G. Wilkerson. 1989. SOYGRO V5.42, Soybean crop growth simulation model. User's guide. Florida Agric. Exp. Stn. Journal no. 8304. IBSNAT (Univ. of Hawaii) and Univ. of Florida, Gainesville.
- Jones, J.W., K.J. Boote, S.S. Jagtap, and J.W. Mishoe. 1991. Soybean development. p. 71-90. *In* Hanks, J., and J.T. Ritchie (ed.) *Modeling plant and soil systems*. ASA, CSSA and SSSA. Agronomy monograph No. 31. Madison, WI
- Jones, J.W., J.W. Hansen, F.S. Royce, and C.D. Messina. 2000. Potential benefits of climate forecasting in agriculture. *Agric. Ecosyst. Environ.* 82:169-184
- Jones, J.W., G. Hoogenboom, C.H. Porter, K.J. Boote, W.D. Batchelor, L.A. Hunt, P.W. Wilkens, U. Singh, A.J. Gijsman, J.T. Ritchie. 2003. The DSSAT cropping system model. *Europ. J. Agronomy* 18:235-265
- Jones, J.W., and J.C. Luyten. 1998. Simulation of biological processes. p.19-62. *In* R.M. Peart and R.B. Curry (ed.) *Agricultural systems modeling and simulation*. Marcel Dekker, New York
- Jun Abe, Donghe Xu, Akane Miyano, Kunihiko Komatsu, Akira Kanazawa, and Yoshiya Shimamoto. 2003. Photoperiod-Insensitive Japanese Soybean Landraces Differ at Two Maturity Loci. *Crop Sci.*, 43:1300-1304

- Kantolic, A.G., and G.A. Slafer. 2001. Photoperiod sensitivity after flowering and seed number determination in indeterminate soybean cultivars. *Field Crops Res.* 72:109-118
- Keating, B.A., P.S. Carberry, G.L. Hammer, M.E. Probert, M.J. Robertson, D. Holzworth, N.I. Huth, J.N.G. Hargreaves, H. Meinke, Z. Hochman, G. McLean, K. Verburg, V. Snow, J.P. Dimes, M. Silburn, E. Wang, S. Brown, K.L. Bristow, S. Asseng, S. Chapman, R.L. McCown, D.M. Freebairn, C.J. Smith. 2003. An overview of APSIM, a model designed for farming systems simulation. *Europ. J. Agronomy*, 18:267-288
- Keim, P., W. Beavis, J. Schupp, and R. Freestone. 1992. Evaluation of soybean RFLP marker diversity in adapted germplasm. *Theor. Appl. Genet.* 85:205-212
- Keim, P., B.W. Diers, T.C. Olson, and R.C. Schoemaker. 1990. RFLP mapping in soybean: association between marker loci and variation in quantitative traits. *Genetics* 126:735-742
- Kennedy, D. 2003. Agriculture and the developing world. *Science* (Washington DC), 302:357
- Kholodenko, B.N., A. Kiyatkin, F.J. Bruggeman, E. Sontag, H.V. Westerhoff, and J.B. Hoek. 2002. Untangling the wires: A strategy to trace functional interactions in signaling and gene networks. *Proc. Natl. Acad. Sci. U.S.A.* 99:12841-12846
- Kirkpatrick, S., C.D. Gelatt, Jr., and M.P. Vecchi. 1983. Optimization by simulated annealing. *Science* (Washington DC), 220:671-680
- Kisha, T.J., B.W. Diers, J.M. Hoyt, and C.H. Sneller. 1998. Genetic diversity among soybean plant introductions and North American germplasm. *Crop Sci.* 38:1669-1680
- Kitano, H. 2002. Computational systems biology. *Nature* (London) 420:206-210
- Knight, J. 2003. A dying breed. *Nature* (London) 421:568-570
- Kobayashi, K., and M.U. Salam. 2000. Comparing simulated and measured values using mean squared deviation and its components. *Agron. J.* 92:345-352
- Kropff, M.J., A.J. Haverkort, P.K. Aggarwal, and P.L. Kooman. 1995. Using systems approaches to design and evaluate ideotypes for specific environments. p. 417-435. *In* Bouma, J. et al. (ed.) *Eco-regional approaches for sustainable land use*. Kluwer Academic Publishers, Dordrecht
- Kumudini, S. 2002. Trials and tribulations: a review of the role of assimilate supply in genetic yield improvement. *Field Crops Res.* 75:211-222

- Kumudini, S., D.J. Hume, and G.Chu. 2001. Genetic improvement in short season soybeans: dry matter accumulation, partitioning and leaf area duration. *Crop Sci.* 41:391-398
- Lark, K.G., K. Chase, F. Adler, L.M. Mansur, and J.H. Orf. 1995. Interactions between quantitative trait loci in soybean in which trait variation at one locus is conditional upon specific allele at another. *Proc. Natl. Acad. Sci. U.S.A.* 92:4656-4660
- Lark, K.G., J. Orf, and L.M. Mansur. 1994. Epistatic expression of quantitative trait loci (QTL) in soybean [*Glycine max*, (L) Merr] determined by QTL association with RFLP alleles. *Theor. Appl. Genet.* 88:486-489
- Leach, J., S. McCouch, T. Slezak, T. Sasaki, and S. Wessler. 2002. Why finishing the rice genome matters. *Science* (Washington DC) 296:45
- Lee, M. 1995. DNA markers and plant breeding programs. *Adv. Agron.* 55:265-344
- Lee, S.H., M.A. Bailey, M.A.R. Mian, T.E. Carter, Jr., D.A. Ashley, R.S. Hussey, W.A. Parrott, and H.R. Boerma. 1996a. Molecular markers associated with soybean plant height, lodging, and maturity across locations. *Crop Sci.* 36:728-735
- Lee, S.H., M. Bailey, M. Mian, E. Shipe, D.A. Ashley, W.A. Parrott, R.S. Hussey, H.R. Boerma. 1996b. Identification of quantitative trait loci for plant height, lodging, and maturity in a soybean population segregating for growth habit. *Theor. Appl. Genet.* 92:516-523
- Leloup, J., and A. Goldbeter. 2000. Modeling the molecular regulatory mechanism of circadian rhythms in *Drosophila*. *BioEssays* 22:84-93
- LeRoy, A.R., W.R. Fehr, and S.R. Cianzio. 1991. Introgression of genes for small seed size from *Glycine soja* into *G. max*. *Crop Sci.*, 31:693-697
- Lockhart, D.J., and E.A. Winzeler. 2000. Genomics, gene expression and DNA arrays. *Nature* (London) 405:827-836
- Loomis R.S., and D.J. Connor. 1992. Crop ecology: productivity and management in agricultural systems. Cambridge University Press, Cambridge
- Maleck, K., A. Levine, T. Eulgem, A. Morgan, J. Schmid, K.A. Lawton, J.L. Dangel, and R.A. Dietrich. 2000. The transcriptome of *Arabidopsis thaliana* during systemic acquired resistance. *Nature Genet.* 26:403-410
- Maloof, J.N., J.O. Borevitz, T. Dabi, J. Lutes, R.B. Nehring, J.L. Redfern, G.T. Trainer, J.M. Wilson, T. Asami, C.C. Berry, D. Weigel, and J. Chory. 2001. Natural Variation in light sensitivity of *Arabidopsis*. *Nature Genet.* 29:441-446

- Manjarrez-Sandoval, R., T.E. Carter, Jr., D.M. Webb, and J.W. Burton. 1997. RFLP genetic similarity estimates and coefficients of parentage as genetic variance predictors for soybean yield. *Crop Sci.* 37:698-703
- Mansur, L.M., K.G. Lark, H. Kross, and A. Oliveira. 1993a. Interval mapping of quantitative trait loci for reproductive, morphological, and seed traits of soybean (*Glycine max*, L.). *Theor. Appl. Genet.* 86:907-913
- Mansur, L.M., J. Orf, K. Chase, T. Jarvik, P. Cregan, and K.G. Lark. 1996. Genetic mapping of agronomic traits using recombinant inbred lines of soybean. *Crop Sci.* 36:1327-1336
- Mansur, L.M., J. Orf, K.G. Lark. 1993b. Determining the linkage of quantitative trait loci to RFLP markers using extreme phenotypes of recombinant inbreds of soybean. *Theor. Appl. Genet.* 86:914-918
- Maughan, P.J., M.A. Saghai Maroof, G. Buss. 1996. Molecular marker analysis of seed weight: genomic locations, gene action, and evidence for orthologous evolution among three legume species. *Theor. Appl. Genet.* 93:574-579
- Mavromatis, T., K.J. Boote, J.W. Jones, A. Irmak, D. Shinde, and G. Hoogenboom. 2001. Developing genetic coefficients for crop simulation models with data from crop performance trials. *Crop Sci.* 41:40-51
- Mavromatis, T., K.J. Boote, J.W. Jones, G.G. Wilkerson, and G. Hoogenboom. 2002. Repeatability of model genetic coefficients derived from soybean performance trials across different states. *Crop Sci.* 42:76-89
- Mayer, D., J. Belward, and K. Burrage. 1996. Use of advanced techniques to optimize a multi-dimensional dairy model. *Agric. Syst.* 50:239-253
- McAdams, H.H., and A. Arkin. 1998. Simulation of prokaryotic genetic circuits. *Annu. Rev. Biophys. Biomol. Struct.* 27:199-224
- McBlain, B.A., and R.L. Bernard. 1987. A new gene affecting the time of flowering and maturity in soybeans. *J. Hered.* 78:160-162
- McBlain, B.A., J.D. Hesketh, and R.L. Bernard. 1987. Genetic effects on reproductive phenology in soybean isolines differing in maturity genes. *Can. J. Plant Sci.* 67:105-116
- Meinke, H., R. Rabbinge, G.L. Hammer, H. van Keulen, and P.D. Jamison. 1998. Improving wheat simulation capabilities in Australia from a cropping systems perspective II. Testing simulation capabilities of wheat growth. *Europ. J. Agronomy* 8:83-99

- Mian, M.A.R., M. Bailey, D.A. Ashley, R. Wells, T. Carter, W. Parrott, H.R. Boerma. 1996. Molecular markers associated with seed weight in two soybean populations. *Theor. Appl. Genet.* 93:1011-1016
- Mian, M.A.R., R. Wells, T. Carter, D. Ashley, H.R. Boerma. 1998. RFLP tagging of QTLs conditioning specific leaf weight and leaf size in soybean. *Theor. Appl. Genet.* 96:354-360
- Morrison, M.J., H.D. Voldeng, and E.R. Cober. 1999. Physiological changes from 58 years of genetic improvement of short-season soybean cultivars in Canada. *Agron. J.* 91:685-689
- Murray, M.G., and W.F. Thompson. 1980. Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res.* 19:4321-4325
- Nelson, R.L. 1986. Defining the seed-filling period in soybeans to predict yield. *Crop Sci.* 26:132-135
- Noble, D. 2002. Modeling the heart—from genes to cells to the whole organ. *Science (Washington DC)* 295:1678-1682
- Oreskes, N., K. Shrader-Frechette, and K. Belitz. 1994. Verification, validation, and confirmation of numerical models in the earth sciences. *Science (Washington DC)* 263:641-646
- Orf, J.H., K. Chase, F.R. Adler, L.M. Mansur, and K.G. Lark. 1999a. Genetics of soybean agronomic traits: II. Interactions between yield quantitative trait loci in soybean. *Crop Sci.* 39:1652-1657
- Orf, J.H., K. Chase, T. Jarvis, L.M. Mansur, P.B. Cregan, F.R. Adler, and K.G. Lark. 1999b. Genetics of soybean agronomic traits: I. Comparison of three related recombinant inbred populations. *Crop Sci.* 39:1642-1651
- Pabico, J.P., G. Hoogenboom, and R.W. McClendon. 1999. Determination of cultivar coefficients of crop models using a genetic algorithm: a conceptual framework. *Trans. ASAE*, 42:223-232
- Pachepsky, L.B., J.D. Haskett, and B. Acock. 1996. An Adequate model of photosynthesis—I Parameterization, validation and comparison of models. *Agric. Syst.* 50:209-225
- Paruelo, J.M., and O.E. Sala. 1993. Effect of global change on maize production in the Argentine Pampas. *Clim. Res.* 3:161-167
- Paterson, A.H., S.D. Tanksley, and M.E. Sorrells. 1991. DNA markers in plant improvement. *Adv. Agron.* 44:39-85

- Perrin, B., L. Ralaivola, A. Mazurie, S. Bottani, J. Mallet, and F. d'Alche-Buc. 2003. Gene networks inference using dynamic Bayesian networks. *Bioinformatics*, 19:138-148
- Pfeiffer, T.W., and D.B. Egli. 1988. Heritability of seed filling period estimates in soybean. *Crop Sci.* 28:921-925.
- Pinheiro, J.C., and D.M. Bates. 2000. *Mixed-Effects Models in S and S-Plus*. Springer-Verlag, New York
- Piper, E.L., K. J. Boote, J.W. Jones, S.S. Grimm. 1996. Comparison of two phenology models for predicting flowering and maturity date of soybean. *Crop Sci.* 36:1606-1614
- Raghothama, C., and A. Pandey. 2003. Absolute systems biology-measuring dynamics of protein modifications. *TRENDS Biotechnol.* 21:467
- Raper, Jr., C.D., and J.F. Thomas. 1978. Photoperiodic alteration of dry matter partitioning and seed yield in soybeans. *Crop Sci.* 18:654-656
- Reymond, M., B. Muller, A. Leonardi, A. Charcosset, and F. Tardieu. 2003. Combining Quantitative Trait Loci Analysis and an Ecophysiological Model to Analyze the Genetic Variability of the Responses of Maize Leaf Growth to Temperature and Water Deficit. *Plant Physiol.* 131:664-675
- Reynolds, J.F., and B. Acock. 1997. Modularity and genericness in plant and ecosystem models. *Ecol. Model.* 94:7-16
- Ritchie, J.T. 1998. Soil water balance and plant water stress. p. 41-54. *In*: Tsuji, G.Y. et al. (ed.) *Understanding options for agricultural production*. Kluwer Academic Publishers, Dordrecht
- Rives, W.A., and T. Galitski. 2003. Modular organization of cellular networks. *Proc. Natl. Acad. Sci. U.S.A.* 100:1128-1133
- Ronald, P., and H. Leung. 2002. The most precious things are not jade and pearls... *Science* (Washington DC) 296:58-59
- Rongwen, J., M.S. Akkaya, and A.A. Bhagwat. 1995. The use of microsatellite DNA markers for soybean genotype identification. *Theor. Appl. Genet.* 90:43-48
- Royce, F.S., J.W. Jones, and J.W. Hansen. 2001. Model-based optimization of crop management for climate forecast application. *Trans. ASAE* 44:1319-1327
- Sachs, K., D. Gifford, T. Jaakkola, P. Sorger, and D.A. Lauffenburger. 2002. Bayesian network approach to cell signaling pathway modeling. *Science STKE* (Available on-line at ([www.stke.org/cgi/content/full/sigtrans;2002/148/pe38](http://www.stke.org/cgi/content/full/sigtrans;2002/148/pe38)); verified 10 Oct. 2003)

- Sandquist, G.M. 1985. Introduction to system science. Prentice Hall, Englewood Cliffs, NJ
- Schenk, P.M., K. Kazan, I. Wilson, J.P. Anderson, T. Richmond, S.C. Somerville, and J.M. Manners. 2000. Coordinated plant defense responses in *Arabidopsis* revealed by microarray analysis. *Proc. Natl. Acad. Sci. U.S.A.* 97:11655-11660
- Schoeberl, B., C. Eicher-Jonsson, E.D. Gilles, and G. Muller. 2002. Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors. *Nature Biotech.* 20:370-375
- Shibles, R., I.C. Anderson, and A.H. Gibson. 1975. Soybean. p. 151-190. *In* Evans, L.T. (ed.) *Crop physiology*. Cambridge University Press, Cambridge
- Sinclair, T.R. 1993. Crop yield potential and fairy tales. p. 707-711. *In* Buxton, D.R. et al. (ed.) *International crop science I*. CSSA publication, Madison, WI
- Sinclair, T.R., S. Kitani, K. Hinson, J. Bruniard, T. Horie. 1991. Soybean flowering date: linear and logistic models based on temperature and photoperiod. *Crop Sci.* 31:786-790
- Smith, J.R., and R.L. Nelson. 1986. Relationship between seed-filling period and yield among soybean breeding lines. *Crop Sci.* 26:469-472
- Smolen, P., D.A. Baxter, and J.H. Byrne. 2000. Mathematical modeling of gene networks. *Neuron* 26:567-580
- Snaidon, G., W.D. Beversdorf, and H.D. Voldeng. 1989. Adjustment of the soybean phenology using the *E4* locus. *Crop Sci.* 29:1361-1365
- Somerville, C., and J. Dangl. 2000. Plant biology in 2010. *Science* (Washington DC) 290:2077-2078
- Somerville, C., and S. Somerville. 1999. Plant functional genomics. *Science* (Washington DC) 285:380-383
- Spaeth, S.C., H.C. Randall, T.R. Sinclair, and J.S. Vendeland. 1984. Stability of soybean harvest index. *Agron. J.* 76:482-486
- Specht, J.E., and J.H. Williams. 1984. Contribution of genetic technology to soybean productivity—retrospect and prospect. p. 49-74. *In* Fehr, W.R. (ed.) *Genetic contributions to yield gains of five major crop plants*. CSSA and ASA Spec. Publ. 7. Madison, WI
- Spirin, V., and L.A. Mirny. 2003. Protein complexes and functional modules in molecular networks. *Proc. Natl. Acad. Sci. U.S.A.* 100:12123-12128

- Stark, J., R. Callard, and M. Hubank. 2003. From the top down: towards a predictive biology of signaling networks. *TRENDS Biotechnol.* 21:290-293
- Staub, J.E., and F.C. Serquen. 1996. Genetic markers, map construction and their application in plant breeding. *HortScience* 31:729-741
- Stewart, D.W., E.R. Cober, and R.L. Bernard. 2003. Modeling genetic effects on the photothermal response of soybean phenological development. *Agron. J.* 95:65-70
- Stoll, M., A.W. Cowley, Jr., P.J. Tonellato, A.S. Greene, M.L. Kaldunski, R.J. Roman, P. Dumas, N.J. Schork, Z. Wang, and H.J. Jacob. 2001. A Genomic-Systems Biology Map for Cardiovascular Function. *Science* (Washington DC), 294:1723-1726
- Strauss, S.H. 2003. Genomics, genetic engineering and domestication of crops. *Science* (Washington DC) 300:61-62
- Stuber, C.W., M. Polacco, M.L. Senior. 2003. Synergy of empirical breeding, marker-assisted selection, and genomics to increase crop yield potential. *Crop Sci.* 39:1571-1583
- Summerfield, R.J., H. Asumadu, R.H. Ellis, and A. Qi. 1998. Characterization of the photoperiodic response of post-flowering development in maturity isolines of soybean [*Glycine max* (L.) Merrill]'Clark'. *Ann. Bot.* 82:765-771
- Summerfield, R.J., R.J. Lawn, A. Qi, R.H. Ellis, E.H. Roberts, P.M. Chay, J.B. Brower, J.L. Rose, S. Shanmugasundaram, S.J. Yeates, and S. Sandover. 1993. Towards the reliable prediction of time to flowering in six annual crops. II Soyabean (*Glycine max*). *Exp. Agric.* 29:253-289
- Swank, J.C., D.B. Egli, and T.W. Pfeiffer. 1987. Seed growth characteristics of soybean genotypes differing in duration of seed fill. *Crop Sci.* 27:85-89
- Tamada, Y., S. Kim, H. Bannai, S. Imoto, K. Tashiro, S. Kuhara, and S. Miyano. 2003. Estimating gene networks from gene expression data by combining Bayesian network model with promoter element detection. *Bioinformatics*, 19:227-236
- Tardieu, F. 2003. Virtual plants: modelling as a tool for the genomics of tolerance to water deficit. *TRENDS Plant Sci.* 8:9-14
- Tasma, I.M., L.L. Lorenzen, D.E. Green, R.C. Shoemaker. 2001. Mapping genetic loci for flowering time, maturity, and photoperiod insensitivity in soybean. *Mol. Breed.* 8: 25-35
- Tasma, I.M., and R.C. Shoemaker. 2003. Mapping flowering time gene homologs in soybean and their association with maturity (*E*) loci. *Crop Sci.* 43: 319-328
- Thomas, J.F., and C.D. Raper, Jr. 1976. Photoperiodic control of seed filling for soybeans. *Crop Sci.* 16:667-672

- Thomas J.F., and C.D. Raper, Jr. 1984. Photoperiod regulation of floral initiation for soybean plants at different ages. *Crop Sci.* 24:611-614
- Thomas, B., and D. Vince-Prue. 1997. *Photoperiodism in Plants*. 2nd ed. Academic Press, New York
- Thompson, J.A., R.L. Nelson, and L.E. Schweitzer. 1995. Relationships among specific leaf weight, photosynthesis rate and seed yield in soybean. *Crop Sci.* 35:1575-1581
- Thornley, J.H.M., and I.R. Johnson. 1990. *Plant and crop modeling*. Clarendon Press, Oxford.
- Thornton, P., and J. MacRobert. 1994. The value of information concerning near-optimal nitrogen fertilizer scheduling. *Agric. Syst.* 45:315-330
- Thum, K.E., D.E. Shasha, L.V. Lejay, and G.M. Coruzzi. 2003. Light- and carbon-signaling pathways. Modeling circuits of interactions. *Plant Phys.* 132:440-452
- Tilman, D., K.G. Cassman, P.A. Matson, R. Naylor, and S. Polansky. 2002. Agricultural sustainability and intensive production practices. *Nature (London)* 418:671-676
- Tischner, T., L. Allphin, K. Chase, J.H. Orf, and K.G. Lark. 2003. Genetics of seed abortion and reproductive traits in soybean. *Crop Sci.* 43:464-473
- Tomking, J.P., and E.R. Shipe. 1996. Soybean growth and agronomic performance in responses to the long-juvenile trait. *Crop Sci.* 36:1144-1149
- Trewavas, A. 2002. Malthus foiled again and again. *Nature (London)* 418:668-670
- Upadhyay, A.P., R.H. Ellis, R.J. Summerfield, E.H. Roberts, and A. Qi. 1994a. Characterization of photothermal flowering responses in maturity isolines of soyabean [*Glycine max* (L.) Merrill] cv. Clark. *Ann. Bot.* 74:87-96
- Upadhyay, A.P., R.J. Summerfield, R.H. Ellis, E.H. Roberts, and A. Qi. 1994b. Variations in the durations of the photoperiod-sensitive and photoperiod-insensitive phases of development to flowering among eight maturity isolines of soyabean [*Glycine max* (L.) Merrill]. *Ann. Bot.* 74:97-101
- Valenzuela, D.M., A.J. Murphy, D. Frendewey, N.W. Gale, A.N. Economides, W. Auerbach, W.T. Poueymirou, N.C. Adams, J. Rojas, J. Yasenchack, R. Chernomorsky, M. Boucher, A.L. Elsasser, L. Esau, J. Zheng, J.A. Griffiths, X. Wang, H. Su, Y. Xue, M.G. Dominguez, I. Noguera, R. Torres, L.E. Macdonald, A.F. Stewart, T.M. DeChiara, and G. Yancopoulos. 2003. High-throughput engineering of the mouse genome coupled with high-resolution expression analysis. *Nature Biotechnol.* 21:652-659
- Vallejos, C.E., N.S. Sakiyama, and C.D. Chase. 1992. A molecular marker-based linkage map of *Phaseolus vulgaris* L. *Genetics* 131:733-740

- van Ittersum, M.K., P.A. Leffelaar, H. van Keulen, M.J. Kropff, L. Bastiaans, and J. Goudriaan. 2003. On approaches and applications of the Wageningen crop models. *Europ. J. Agronomy* 18:201-234
- van Schaick, P.H., and A.H. Probst. 1958. Effects of some environmental factors on flower production and reproductive efficiency in soybeans. *Agron. J.* 50:192-197
- Venables, W.N., and B.D. Ripley. 1997. *Modern applied statistics with S-Plus*. 2nd. ed. Springer-Verlag, New York
- Voldeng, H.D., E.R. Cober, D.J. Hume, C. Gillard, and M.J. Morrison. 1997. Fifty-eight years of genetic improvement of short-season soybean cultivars in Canada. *Crop Sci.* 37:428-431
- Wardlaw, I.F. 1990. The control of carbon partitioning in plants. *New Phytol.* 116:341-381
- Weckwerth, W. 2003. Metabolomics in systems biology. *Annu. Rev. Plant. Biol.* 54:669-689
- Welch, S.M., J.L. Roe, and D. Zhanshan. 2003. A genetic neural network model of flowering time control in *Arabidopsis thaliana*. *Agron. J.* 95:71-81
- Wells, R., L.L. Schulze, D.A. Ashley, H.R. Boerma, and R.H. Brown. 1982. Cultivar differences in canopy apparent photosynthesis and their relationship to seed yield in soybeans. *Crop Sci.* 22:886-890
- Weng, G., U.S. Bhalla, and R. Iyengar. 1999. Complexity in biological signaling systems. *Science (Washington DC)* 284:92-96
- White, J.W. 1998. Modeling and crop improvement. p. 179-188. *In* Tsuji, G.Y. et al. (ed.) *Understanding options for agricultural production*. Kluwer Academic Publishers, Dordrecht
- White, J.W., and G. Hoogenboom. 1996. Simulating effects of genes for physiological traits in a process-oriented crop model. *Agron. J.* 88:416-422
- White, J.W., and G. Hoogenboom. 2003. Gene-based approaches to crop simulation: past experiences and future opportunities. *Agron. J.* 95:52-64
- Wilcox, J.A., W.J. Wiebold, T.L. Niblack, and K.D. Kephart. 1995. Growth and development of soybean isolines that differ for maturity. *Agron. J.* 87:932-935
- Wilcox, J.R., W.T. Schapaugh, Jr., R.L. Bernard, R.L. Cooper, W.R. Fehr, and M.H. Niehaus. 1979. Genetic improvement of soybeans in the Midwest. *Crop Sci.* 19:803-805
- Wonder wheat*. 1998. *Science (Washington DC)* 280:527

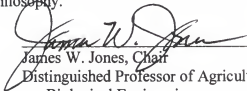
- Yamanaka, N., S. Ninomiya, M. Hoshi, Y. Tsubokura, M. Yano, Y. Nagamura, T. Sasaki, and K. Harada. 2001. An informative linkage map of soybean reveals QTLs for flowering time, leaflet morphology and regions of segregation distortion. *DNA Res.* 8:61-72
- Yin, X., M.J. Kropff, and P. Stam. 1999. The role of ecophysiological models in QTL analysis: the example of specific leaf area in barley. *Heredity* 82:415-421
- Yin, X., P. Stam, M.J. Kropff, and A.H.C.M. Schapendonk. 2003. Crop modeling, QTL mapping, and their complementary role in plant breeding. *Agron. J.* 95:90-98
- Yu, J., S. Hu, J. Wang, G.K. Wong, S. Li, B. Liu, Y. Deng, L. Dai, Y. Zhou, X. Zhang, M. Cao, J. Liu, J. Sun, J. Tang, Y. Chen, X. Huang, W. Lin, C. Ye, W. Tong, L. Cong, J. Geng, Y. Han, L. Li, W. Li, G. Hu, X. Huang, W. Li, J. Li, Z. Liu, L. Li, J. Liu, Q. Qi, J. Liu, L. Li, T. Li, X. Wang, H. Lu, T. Wu, M. Zhu, P. Ni, H. Han, W. Dong, X. Ren, X. Feng, P. Cui, X. Li, H. Wang, X. Xu, W. Zhai, Z. Xu, J. Zhang, S. He, J. Zhang, J. Xu, K. Zhang, X. Zheng, J. Dong, W. Zheng, L. Tao, J. Ye, J. Tan, X. Ren, X. Chen, J. He, D. Liu, W. Tian, C. Tian, H. Xia, Q. Bao, G. Li, H. Gao, T. Cao, J. Wang, W. Zhao, P. Li, W. Chen, X. Wang, Y. Zhang, J. Hu, J. Wang, S. Liu, J. Yang, G. Zhang, Y. Xiong, Z. Li, L. Mao, C. Zhou, Z. Zhu, R. Chen, B. Hao, W. Zheng, S. Chen, W. Guo, G. Li, S. Liu, M. Tao, J. Wang, L. Zhu, L. Yuan, and H. Yang. 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science* (Washington DC) 296:79-91.
- Zhang, L., R. Wang, and J.D. Hesketh. 2001. Effects of photoperiod on growth and development of soybean floral bud in different maturity. *Agron. J.* 93:944-948
- Zhu, Y.L., Q.J. Song, D.L. Hyten, C.P. Van Tassell, L.K. Matukumalli, D.R. Grimm, S.M. Hyatt, E.W. Fickus, N.D. Young, and P.B. Cregan. 2003. Single-nucleotide polymorphisms in soybean. *Genetics* 163:1123-1134

## BIOGRAPHICAL SKETCH


Carlos D. Messina was born and raised in the suburbs of Buenos Aires, Argentina. He received a Bachelor of Science degree in agronomy and a master's degree in crop production from the University of Buenos Aires, followed by a Ph.D. in agricultural and biological engineering from the University of Florida. He was a research assistant in the IFEVA (Institute of Physiological and Ecological Research in Agriculture) at the University of Buenos Aires, and in the Crop Systems Modeling Laboratory at the University of Florida. His research interests include plant systems biology, bean genomics, simulation and optimization of food production systems, and climate forecast applications in agriculture. Carlos has published in *Agricultural Systems*, *Agricultural and Forest Meteorology*, and the *Journal of Applied Meteorology*.

Carlos has received several awards. In 1998 he was the recipient of a fellowship from the IAI (Inter-American Institute for Global Change Research) for collaborative research in climate forecast applications research at the University of Miami and the University of Florida. He previously received a two-year scholarship for his MS program and a one-year fellowship granted by AACREA (Argentina Association of Agricultural Research Consortia). He is listed in *Who's Who in American Universities and Colleges*, and has been inducted into Gamma Sigma Delta, the Honor Society of Agriculture.

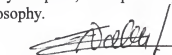
I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

  
James W. Jones, Chair  
Distinguished Professor of Agricultural and  
Biological Engineering

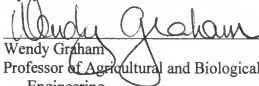
I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

  
Kenneth J. Boote, Cochair  
Professor of Agronomy


I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

  
C. Eduardo Vallejos  
Associate Professor of Horticultural  
Sciences

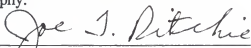
I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

  
Wendy Graham  
Professor of Agricultural and Biological  
Engineering

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

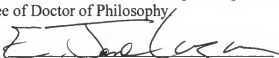
  
Bernard A. Hauser  
Assistant Professor of Botany

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

  
\_\_\_\_\_  
Joe T. Ritchie  
Distinguished Professor of Agricultural and  
Biological Engineering

This dissertation was submitted to the Graduate Faculty of the College of Agricultural and Life Sciences and to the Graduate School and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy

December of 2003

  
\_\_\_\_\_  
Dean, College of Agricultural and Life  
Sciences

\_\_\_\_\_  
Dean, Graduate School